IRE Transactions

on Microwave Theory and Techniques

Volume MTT-9 **SEPTEMBER, 1961** Number 5

Published Bimonthly

In This Issue

PUBLISHED BY THE

essional Group on Microwave Theory and Techniques

Microwaves—A Review of Progress in Great Britain During 1960*

A. E. KARBOWIAK†

Summary—**The fields covered by the review are: Electromagnetic theory as applicable to wave propagation, theory of waveguides and components, microwaves in fundamental measurements (such as time and atomic constants), designs of components, measurements, solid state devices and a summary of utilization. Future trends are also indicated. The review (with some exceptions) does not cover the following fields: Antennas, propagation and microwave tubes. In reviewing technical publications, preference is given to such aspects of work which are novel, fundamental or are of controversial nature or are likely to influence future trends; where appropriate, criticism of the work is given.**

Introduction

MICROWAVES is a term (now almost generally accepted) which stands for a vast field of activity. Indeed, so vast that many of us who have been active in it for many years have been surprised by its tremendous rate of growth. Not surprisingly, therefore, one finds that a substantial proportion of published scientific papers, public lectures on science and various aspects of engineering (not only electrical) as well as subjects on utilization, directly or indirectly concern microwaves: microwaves *per se*, and microwaves for fundamental measurements on atomic and material constants. The most precise measurements of length and time involve microwaves and *vice versa*, for some microwave uses, the skill of the mechanical engineer is tested to the full. Microwave power from fractions of a microwatt to many megawatts and frequencies corresponding to wavelengths from decimeters to fractions of a millimeter is now a common practice.

Industries of diverse interests call on microwaves as the new source of power since microwaves, when liberated in the form of heat, can be applied with precision in the quantity and place required. Greater and greater powers also become available for communication purposes and this brings an investigation of industrial, biological as well as medical aspects of microwave radiation into prominence.

Parametric amplifiers and masers are the fruits of research carried out during the last decade and they are often applied in the fields of radio astronomy and communications. In fact it is said that there has never been so much interest in liquid helium as there has been since the advent of low noise receivers!

For applications involving microwaves in communication systems, as well as in fundamental research, there is a trend towards higher and higher frequencies, yet lack of suitable sources of power in the millimetric and submillimetric bands besets any rapid progress. Microwave tubes are inefficient, have short life and are not a commercial proposition in the short millimetric band. But, for submillimetric frequencies harmonic generators are the only sources of monochromatic power. These are inefficient and progress is slow and doubtful. Not surprisingly, new solid-state devices of unspecified nature are the hope; the stakes are high but odds are equally high. This is a field for fundamental research and so is the whole field of generation of coherent radiation in the band from submillimetric waves through infrared to frequencies corresponding to the visible spectrum. Are we going to include this entire field under the auspices of microwaves, particularly since our many measurement methods of the higher microwave frequencies already resemble those of optics?

It is fitting to observe that among the many conventions, conferences and international meetings which have taken place in Great Britain during 1960, three in particular were concerned, within their own spheres of interest to a larger or smaller extent, with microwaves. These were:

1) The XIIIth General Assembly of URSI held at University College, London, Eng., September 5–15, 1960.
2) A Conference on Solid State Microwave Amplifiers held at Nottingham University, Eng., April 6–8, 1960.
3) The 3rd International Conference on Medical Electronics held at Olympia, London, July 21–27, 1960, under the auspices of the IEE and the International Federation for Medical Electronics.

Work in the microwave field can conveniently be reviewed under the following headings:

1) Investigation of fundamentals of physics such as measurements of length and time and atomic and material constants.
2) Investigation of fundamentals of electromagnetic waves.
3) Development in microwave components and measurements.
4) Generation of microwave radiation, in particular solid-state devices.
5) Propagation and antennas.
6) Utilization including industrial applications and communication systems.
7) Future trends.

* Received by the PGMTT, June 2, 1961.
† Standard Telecommunication Labs., Ltd., Harlow, Essex, Eng.

The above topics cover an immense territory of our knowledge of which only a part was chosen for this review.

Under item 1 a great deal of work done in physics, such as spectroscopy, could be listed but because of its very specialized nature it will be omitted from further discussion. On the other hand the measurements of length and time deserve a mention, because such measurements are fundamental.

Study of microwaves and waveguides has always involved a great deal of electromagnetic theory and not infrequently progress in the field of microwaves was a direct result of deeper understanding of electromagnetic theory; a step forward in the electromagnetic theory is always a step forward in the field of microwaves.

Progress in microwave components and measurements is usually of the nature of an improvement or extension in the existing knowledge and consequently no detailed discussion of such work will be given here. Nevertheless, such contributions are building-blocks of microwave techniques and some references are therefore included.

Any progress in the generation of microwave power, particularly in the upper frequency band is of utmost significance in further progress in the microwave field; but progress in conventional microwave tubes has been extensively reviewed recently by Harvey [1]. This review will be chiefly concerned with generation of microwaves using solid-state devices, a fascinating field full of vague promises.

Propagation and antennas is a subject intimately related to the activities of URSI and progress in this field can be gathered from the many excellent reports of the XIIIth General Assembly, which took place in the autumn of 1960. Mention should be made however of the work by Belatini [2] which does not seem to have received any publicity. The work which concerns tropospheric scatter phenomena, deserves attention because there exists a considerable disagreement over the exact process of the scattering phenomenon. Belatini observes that the field received beyond the horizon shows an almost regular spatial interference pattern, implying that incoherent scatter cannot be the mechanism responsible, as is so widely accepted, for signals received beyond the horizon. The author's thesis is that the heat from the sun is responsible for vertical motion of air producing in effect torus-shaped "Bénard cells." These cells act as weak biconcave lenses causing the radio waves to diverge in the vertical plane and in this way enable radio signals to reach beyond the horizon. The hypothesis is not only original, but the evidence against "tropospheric scatter" is strong and the evidence for Belatini's theory good.

Industrial applications, and communication systems in particular, are the important uses for microwave equipment. Much of the interest and financing of research in industries is taking place because of such applications. The existence of research, therefore, hinges to a large extent on possible industrial and commercial applications although research *per se* is being accepted on a larger and larger scale. Microwave links are, to a large extent, the main commercial interest, but, in addition, numerous defence projects such as radar, etc., occupy the activity of many. But new communication systems such as long distance communication by waveguide and satellites occupy the imagination of many. Some other industrial applications have been discussed adequately by Harvey [3].

Assessment of future trends, which is of great importance to industry and those responsible for planning future activities, involves a specific form of crystal-gazing. But, there is one certain lesson that we can learn from the past: few things in life are less predictable than the activity of a research worker; in fact the game is so speculative that staking a Derby winner is not an unjust analogy!

This review is primarily concerned with work which in the opinion of the reviewer is either of a fundamental nature or is likely to have a significant impact on the progress in microwaves, has features of novelty or is simply fascinating.

Fundamentals

A concise review of progress in electrical units and standards has been prepared by Vigoureux [4]. It transpires that the fundamental units of electricity and magnetism can now be measured with a reproducibility of about 1 part in 10^6 using electrical measurements on constants such as gyromagnetic ratio of the proton.

Length and time—the two fundamental units—can be measured and the units maintained and reproduced to a much higher order of accuracy than was possible only a few years ago. The measurements are carried out in electromagnetic spectrum; but whereas the "standard of length" is conveniently found in the visible spectrum (for example, the wavelength of the orange line of krypton—86), the most accurate "clock" operates at a fundamental frequency falling in the microwave band. The caesium frequency standard which operates at about 9.2 Gc, is a result of three or four years of research carried out in various countries.

In Great Britain there are now a number of caesium clocks in operation following the fundamental work carried out by Essen and Parry at the National Physical Laboratory [5]. Such clocks operate with an error not exceeding about 1 part in 10^{10}.

In March, 1959, the International Committee of Weights and Measures adopted the second of Ephemeris Time as the official standard. The astronomical standard, however, cannot be reproduced to better than about 1 part in 10^9, even when averaged over a period of several years. The caesium beam standard, on the other hand, is readily available and there is a possibility that the second of time may be defined in terms of frequency

of an atomic frequency, and this lies in the microwave band.[1]

Quartz crystal oscillators are still a very convenient intermediary between the caesium clock and astronomical observations. But with the recent developments in the atomic frequency standards and also MSF transmissions in the United Kingdom and Western Europe, the quartz crystal oscillators are likely to become superfluous.

Present work on atomic frequency standards is concerned chiefly with design of more compact units (some so small that they can be fitted inside rockets or installed inside an artificial satellite), study of different designs, sources of errors and intercomparison of measurements. There is at present an unresolved discrepancy of about 2 parts in 10^{10} between clocks of different type or build in different countries [6].

A characteristic feature of many meetings or lectures concerning microwaves is that the subsequent discussions are frequently diverted towards the fundamentals of electromagnetic theory. There again one is frequently involved in discussion on basic units [7], or even the choice of coordinate systems and nomenclature. The opinions are frequently diverse and sometimes no conclusions are reached. This is all to the good, yet nevertheless a little surprising, for it is an outward sign of an inward uneasy feeling about the fundamental concepts. Most of us, at one time or another, when engaged in an investigation of an advanced nature have experienced a feeling of complete standstill, only to find out that the root of the trouble was inadequate comprehension of fundamentals and not the apparent complexity of the particular problem on hand. We should, rightly so, have a look at our fundamental formulas and examine them from time to time in the light of gathered experience. It is reassuring to find that this spirit is prevailing and our concern about the fundamental concepts is a healthy one.

The Department of Scientific and Industrial Research has issued a booklet containing tables [8] of the refractive index of air for a wide range of temperatures, pressures and water-vapor pressures based on formulas derived by Essen and Froome.

Theory of Wave Propagation and Waveguides

The foundations of waveguide theory were laid down more than half a century ago by Lord Rayleigh, but it was the widespread adoption of the powerful analytical techniques from the field of quantum mechanics directly to microwaves that led to the elaboration of waveguide theory as we know it today. In principle, therefore, any imaginable wave-guiding structure could be analyzed in detail, yet it would be a gross overstatement to say that there is fundamentally nothing new that could be discovered. For there is more to understanding than mere writing down of general mathematical expressions. One must be able to comprehend what the mathematical expressions stand for and be able to formulate useful physical pictures before a clear physical interpretation can be given; and later on one must be able to apply it to the design of a particular structure, be it a single component or a complete communication system. It is clear that during the last few years we have progressed towards that goal, but we have not reached it entirely. Indeed, it is doubtful if such an ideal state of understanding can ever be reached. But, it is only through discussion, analysis, and a continual examination and re-examination of our accumulated facts that a progress towards unification of our knowledge can be made. At the same time we may be able to derive some enjoyment from it.

There are two outstanding aspects of the theory of waveguides: 1) propagation in waveguides with walls other than perfectly conducting, and 2) propagation in nonuniform waveguides.

Propagation in Homogeneous Uniform Waveguides

The two above mentioned factors have one thing in common: we do not seem to be entirely sure what the best way is to describe an electromagnetic field inside a waveguide and how such fields should be classified. A particular case is that of surface waves.[2]

Some contributors, in their own way, maintain that a rational unified and useful classification of waveguide modes is possible. Thus, Karbowiak [9] shows quite generally that the electromagnetic field can be conveniently resolved into pole waves and branch-cut waves; but in addition he also introduces the concept of a quasi-mode. The findings can be briefly summarized as follows.

To every cylindrical waveguide, there corresponds [10] a certain infinite but discrete distribution of poles in the complex plane of the propagation coefficient, γ, and one or two branch points (usually only one) depending on the number of geometrically possible infinite directions, in the transverse plane of the waveguide. Integration around the branch-cut yields the branch-cut wave, and the poles represent the possible modal (or guided wave) propagation. This picture is agreed on by most workers in the field. Karbowiak however accepts for the definition of a pole wave the following function: $\psi_i = a_i \cdot f_i(u, v) \cdot \Lambda_i \cdot \exp(-\gamma_i z)$, where a_i is the amplitude coefficient, $f_i(u, v)$ a function describing the field pattern in the waveguide transverse plane and Λ_i a multiplier, which is related to the

[1] The implied great faith in the stability of atomic vibrations, which we seem to cherish, cannot be overemphasized, particularly since our knowledge in connection with atoms and molecules is still so new.

[2] It was the function of a Study Group of Commission VI of URSI to try to discuss various types of surface waves with a view to classifying them. But, it transpired that there are nowadays so many variants of surface wave concepts that a useful definition could not be formulated. It was agreed that to pursue the matter of terminology applied to surface waves would be unprofitable.

complex Fresnel Integral. This definition differs from the more commonly accepted one by the multiplier Λ. If $\Lambda = 1$, then the expression defines a quasi-mode. The point of such a definition will transpire from an example.

Imagine a waveguide with a given geometry and excitation having arbitrary but physically realizable walls. Then in the plane of γ there will be a distribution of an infinity (discrete) of poles and, say, one branch point. Of the poles, some will be placed on the right and some on the wrong Riemannian Plane [10]. The quantity Λ will be close to unity for all poles on the "right" plane and it will be very nearly zero for poles lying on the "wrong" plane. But, Λ is a continuous function of frequency and parameters describing the physical medium and, therefore, if either of these quantities is gradually varied, then the poles will move from place to place and the value of the multiplier will vary. Some poles, however, will pass gradually through the branch cut onto the other Riemannian plane, but the total number of poles will remain unaltered, and, therefore, all pole waves exist provided the excitation is such that a_i is finite. The constitution of the field changes continuously, because depending on the value of the multiplier Λ some pole waves become more significant than others. If we were to expand the field in terms of functions with $\Lambda = 1$ then clearly the modes would either exist or not depending on whether the pole were just to the left or the right of the branch cut. Since all events in nature are continuous such a classification could not be recommended. Yet, functions with $\Lambda = 1$ are useful in describing fields in finite regions and for such modes the term quasi-mode is reserved; modes of this kind need not, on their own, be a solution to the waveguide problem, but in *finite* regions (of interest) the field may very closely approximate such modes and hence their significance. A leaky wave is one example, a Zenneck wave another.

Some investigators, however, do not share Karbowiak's viewpoint. Barlow [11] for example, proposes to classify electromagnetic waves into three types: 1) TEM waves, 2) surface waves and 3) so-called wave guide modes. Such classification, while having the merit of apparent simplicity, would be rather restricted, as pointed out recently by Waldron [12]. Because, if we consider the pole representing (say) the TEM wave in a perfect waveguide, then as we change gradually the perfectly conducting medium and make it imperfect, the pole will move slowly away from the imaginary axis and the associated Λ will become a little less than unity. Eventually, if the properties of the media are sufficiently altered, then the mode will not in the least resemble the TEM mode. The wave due to that particular pole may even be more like some other mode or even form a hybrid with one of its neighbors, etc. As a pole wave, however, it exists all the time, though for some properties of the media it may appear as one quasi-mode and for some as another.

Surface waves in various forms have always been a fruitful field for research—surface waves for transmission lines or for antennas, surface waves for measurements on materials and surface waves for their own sake.

Clarricoats has made an interesting and exhaustive study of wave propagation along dielectric rods [13]. In the past, several investigators have analyzed propagation of selected modes on dielectric cylinders with and without a concentric metallic shield. In this study, however, the author was able, through a detailed analysis, to show how the various modes (which sometimes are very complicated) are related to each other and how the mode spectrum is affected by the cross-sectional dimension of the rod and the proportions of the metallic shield. This is yet another example of a unification of the concept of modes and a greater understanding of the structure of electromagnetic field. A graphical method of evaluation of propagation coefficients is given and application of the results to ferrite devices is also indicated.

The particularly interesting feature of the analysis is the physical picture. The mode spectrum of the structure can be followed step by step as the diameter of the dielectric rod and the diameter of the metallic shield are varied. In fact, the modes of an empty cylindrical waveguide are linked to those of a shielded dielectric rod and these in turn are related to those of an unbounded dielectric rod, the surface wave modes. It would be interesting to analyze the excitation problem in such waveguides. From such an analysis, we venture to guess, it would transpire that there is a branch-cut wave representing the radiation field, and a discrete infinity of pole waves. The pole waves would always be present but their proportion in the total field and their character would vary as the proportions of the dielectric rod and the metallic shield were varied. In this way modes of one waveguide could be related to modes in another waveguide by a continuous process without any ambiguity. Furthermore, over certain finite regions some of the modes (quasi-modes) would be appreciable and some other modes would be negligible (the quasi-mode would have no physical significance).

A number of other surface wave modes and supporting structures have also been analyzed. The hybrid EH_{11} mode of a dielectric rod has received detailed attention with a view to utilizing this mode to measurements of permittivity of samples in the form of circular rods [14]. Surface wave antennas have been analyzed on many occasions previously but Hirsch [15] has carried out an analysis of higher-order surface wave modes supported on a dielectric coated conductor. These modes are all hybrid modes and most of them are very complicated. But although such modes can have low attenuation, on suitably proportioned guides, the exploitation is difficult because of analytical complexity and lengthy computations.

Two different surface wave excitation problems have

been analyzed. One by Wait and Conda [16] concerns essentially launching efficiency of a surface wave from a magnetic line source placed above a reactive surface. The novel part of the investigation lies in the extension of the analysis to launching of a surface wave over a corrugated elliptic cylinder. It is shown that for certain proportions of the cylinder strong resonance characteristics are obtained. The results, no doubt, can be extended to other structures such as radiation over the surface of a sphere. Another excitation problem was analyzed by Angulo and Chang [17]. This concerns excitation of surface waves along two infinite, identical, separate and parallel dielectric slabs by a magnetic line source. A modal analysis is carried out and an expression for the far field is obtained. Launching efficiency, radiation loss and radiation pattern are also obtained. It is shown that there exists an optimum thickness and spacing of the slabs for maximum efficiency. A surface wave problem of a different kind was analyzed by Williams [18]. He investigates the excitation of surface waves by plane electromagnetic wave incident on a wedge with surface impedancy boundary conditions. It is shown that surface waves will be excited even though the surface impedance values on the two sides of the wedge are unequal, provided that certain discrete values of the wedge angle are excluded.

There has been, and still is, a considerable amount of interest and discussions on how surface waves should be classified. But, in the opinion of the reviewer an important aspect is still not receiving adequate attention, that is, we are calling by the same name "surface wave" physically different things.[3] On one side there are the modal waves (pole waves) of propagation and these could be classified on their own. On the other side there are various types of waves associated with the radio wave propagation over the ground. We do not think that it really matters whether we call the first group surface waves and call the other group by a different name or the other way round. Provided that a distinction between the two groups (let it be three groups, if the need arises) is made, classification within the group is of secondary importance and will follow much more easily. After all, biologists, for example, do not call different species of the animal kingdom by the same name, but find it easier to differentiate between them, despite the fact that borderline cases clearly exist.

Propagation in Nonuniform Waveguides

There are many variants of nonuniform waveguides and these are discussed under three separate headings: 1) Single mode propagation in nonuniform lines, 2) Propagation in slowly varying nonuniform multimode waveguides and 3) Propagation in waveguides with statistically distributed irregularities.

[3] A good summary of the situation is given in a review paper by J. R. Wait presented at the XIIIth General Assembly of URSI.

The foundations of the theory of wave propagation in nonuniform waveguides were established by many workers in different countries (*e.g.*, Stevenson, Schelkunoff, Katsenelenbaum, Riter, etc.). The subject is very complicated, but by making various simplifying assumptions it is possible to solve a number of specific problems and to go a long way towards forming a coherent physical picture.

Problems involving only a single mode can be reduced to the problem of design of a taper between two transmission lines of unequal characteristic impedance. It is well known that Tchebycheff design leads to minimum length for a given reflection coefficient over an infinite bandwidth. But, for most applications only a finite bandwidth is required and then it is possible to design a shorter taper. Solymar shows that there is an infinity of possible solutions to tapers of as short a length as desired but that the shape of the taper would be extremely complicated (a violently oscillating function), and therefore the design would be impractical [19]. By introducing the concept of "function complexity" he was able to limit the class of functions and give design criteria of practical significance. The theoretical results are interesting in that there exists a Fourier transform relation between the shape of the taper and the reflection coefficient.

A problem somewhat related to tapers is that of bends. Here again over a given bandwidth a small reflection coefficient is specified and the problem is to design a bend as small as possible. The usual procedure is to design a binomial bend; but Wray and Hastie show that by going to the limit and taking infinitesimal sections, a considerable improvement in the design is possible [20]. The analogy with taper is striking, and one would guess that further improvement should be possible using Solymar's approach. Presumably many other gradually varying transitions could be treated in a similar manner, *e.g.*, twisted waveguides [21].

An entirely different approach to tapers was proposed by Barlow [22]. The idea is to maintain a completely undisturbed field pattern not only inside but also outside the guide, irrespective of its cross section. The method has been demonstrated by application to the H_{0n} and E_{0n} modes but it is doubtful whether it will be possible to extend it to other modes because physically it would be difficult to provide the required boundary conditions. We can examine the proposal by reference to Fig. 1. Here A and B are arbitrary planes extending right across the infinite pattern of the H_{0n} (or E_{0n}) mode. The planes A and B could then be replaced by suitable impedance sheets and the field outside omitted. Clearly, A and B form, geometrically, a taper, but not from an accepted electrical point of view. The reason is that the wave is not really "guided" since the energy crossing planes A and B is of the same order of magnitude as the component of energy flow along the planes. The wave pattern does remain undisturbed but the planes A and B are not surface im-

pedance boundaries and they do not "guide" the wave. No doubt, some applications of this proposal will be forthcoming and any experimental work in this field would be of interest.

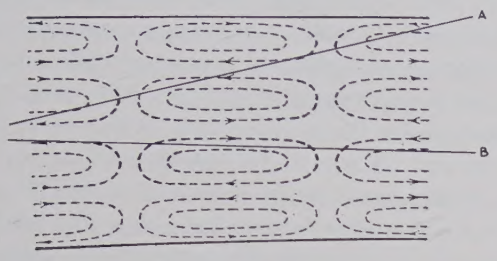


Fig. 1.

Design of mode transducers is perhaps one of the most difficult problems in the theory of nonuniform waveguides and has eluded a satisfactory solution for a long time. Such designs as have been produced have been based on engineering intuition. Solymar and Eaglesfield have recently proposed a new approach to the design of mode transducers [23]. In essence the method of design is as follows. If a mode ψ_1^A in waveguide A is to be transformed into a mode ψ_2^B in waveguide B then we form a wave function $\psi = g_1(z)\psi_1^A + g_2(z)\psi_2^B$ where g_1 and g_2 are slowly varying functions of the axial coordinate z with terminal conditions $g_1(0) = g_2(L) = 1$ and $g_2(0) = g_1(L) = 0$, where L is the length of the transducer. There is, of course, an infinite number of possible solutions and the optimum solution is yet to be discovered. For simplicity the eigenvalues of A and B are made equal by preceding either one or the other waveguide with a simple taper. The shape of the transducer is determined by zeros of the ψ function; this part of the design for all but the simplest cases is best accomplished by a graphical procedure. The authors have designed a number of different transducers and the simplest example is that of the H_{01} rectangular to H_{02} rectangular [24]. There does not seem to be any limitation to the method, and transducers so designed are all wide band. It is most instructive to work out a complete design of a transducer because it is then that one gains a deeper and clearer understanding of modes in nonuniform waveguides.

It remains to be proven,[4] but it would appear that the complete set of pole waves is undergoing a continuous transformation as one continuously changes the shape of the waveguide. There is a one-to-one correspondence between modes in one section of the waveguide and another. No new poles can be created by such a continuous process but some poles may merge into one or form hybrids, while others, which may have been presupposed to be simple poles, will split. It is not possible to subdivide modes in general waveguides into E or H modes or in any other way that would be generally valid, but one can label the modes just the same. During such a process one should not be surprised to find that a pole representing an E_{01} mode in a circular waveguide will become transformed by a one-to-one correspondence into (say) an H_{27} mode of a circular waveguide, but, if the transformation of the waveguide is carried out in a different way, a one-to-one correspondence with another mode will be established. The type of one-to-one correspondence between the modes is purely a function of the shape of the transducer.

Transmission properties of waveguides with statistically distributed irregularities are of particular importance in connection with long distance communications by waveguide. Various analyses have been carried out in Great Britain and the United States during the last few years, but recently Larsen (in Germany) has prepared yet another study [25].

Waveguide Components, Devices and Measurements

The field of interest here can be divided into theoretical analysis of components and various configurations of obstacles and apertures, properties of ferrites and measurements of dielectric properties and microwave power.

Strip-lines continue to be a subject of interest [26], [27]. With most strip-line configurations, it is difficult to obtain a good correlation between theory and experiment. This is understandable because the structure, from an electrical point of view, is rather complicated. In addition, strip-line forms an open structure and, therefore, any discontinuities such as matching stub or corner cannot be calculated using conventional waveguide methods because of the associated radiation field. Some progress, however, can be made using various simplifying assumptions, but the results are understandably approximate. Assuming a substantially TEM transmission, Lewin was able to assess the properties of a post, a corner and a short circuit; he also examined the radiation patterns associated with such discontinuities [28]. He concluded that a short circuit is appreciably worse, from the point of view of radiation loss, than an open circuit. Only a few of his results, however, have received experimental confirmation to date.

The theory of waveguide junctions is now well understood through Schwinger's Variational Approach. But all waveguide discontinuities could be analyzed from first principles by matching the complete waveguide fields on both sides of the discontinuity. The matrix relating the modes is infinite, but for a single mode waveguide the elements of the matrix—which, with the exception of the dominant mode, represent only evanescent modes—is quickly convergent. Sinha used this fact and approximated by summing up just the first few terms and in this way was able to obtain the

[4] The description given in this paragraph represents the reviewer's opinion. No rigorous proof in support of it has yet been produced.

equivalent circuit parameters for a waveguide junction [29]. His formulas agree very well with experimental results.

A problem of great practical significance and yet difficult to analyze for the general case is coupling of waveguides through apertures. Some years ago Stevenson obtained a rigorous formulation of the problem, but his equations are difficult to apply in most cases. Bethe's approach is relatively simple and has been used successfully for the design of many waveguide and cavity configurations. Bethe's method, however, is restricted to small apertures only, and a new approach, to handle the numerous practical cases where large apertures are used, is clearly needed. Lewin's approach forms a useful extension to Bethe's work [30]. The idea is to use the results obtained in the antenna theory (quasi-static) and apply them *via* Babinet's principle to apertures in a waveguide wall and obtain corrections to Bethe's aperture polarizabilities. It would be difficult to deal, in this way, with arbitrary apertures but the author manages to analyze long narrow slots and also a case of two waveguides coupled by crossed slots.

The phenomenon of traveling-wave resonance has a number of specialized applications. The theory and some of its properties were investigated by Twisleton [31].

Theory of microwave propagation in ferrites is of great importance in the design of numerous microwave components such as isolators, gyrators, phase shifters, etc. But, because of the tensorial character of ferrite permeability, the general theory is crowded with difficulties. Although in a number of specific cases, useful solutions can be obtained [32]. Sometimes, even what could appear, on superficial examination, to be quite a manageable problem, seems to present difficulties. Thus, for example, the determination of the input impedance of a ferrite-filled waveguide becomes a sizable analytical problem to which only partial answers have been obtained. For certain range of parameters Lewin was able to obtain a correct solution, but outside that region the formulas lead to a physically unacceptable solution (complex impedance) and have to be rejected [33]. Recently however, by taking into account Bresler's surface waves, which propagate in the gap between the ferrite and the wall, the author was able to justify the validity of his solutions.

Various measurements on ferrites and their interpretation are yet another prolific source of mathematical problems. One of the most favored measurement methods is one in which the sample of ferrite to be measured is prepared in the form of a small sphere and is inserted inside a cavity [34]. From the measurement of the change in the resonance conditions of the cavity the ferrite permeabilities are determined. In the past a number of perturbational techniques were developed, and there have been many discussions concerning the relative merits of various methods. Underlying all theoretical formulas there are various assumptions and these have been reviewed and classified by Waldron [35].

Various measurements on waveguide components, power measurements using resistive films and Q-factor measurements have also been described [36]–[40]. These incorporate various extensions and modifications to the existing methods. A somewhat different method of measuring permittivity has been proposed by Sinha and Brown [41]. Here the sample, which is in the form of a thin rod, is inserted through a hole in the end wall of a circular cavitiy operated in the H_{01n} mode. The other end wall is formed by a movable plunger. At the desired frequency the loss tangent and permittivity are deduced, from the cavity Q factor and the change in the plunger position, as a function of the sample insertion.

In the millimetric band, yet another method of measuring dielectric properties becomes feasible; Caicoya refers to the method as the "High-Order Mode Interferometer" [42]. In essence, the method consists of transmitting a parallel beam of millimetric radiation at an angle to a parallel slab of material (under test) placed between metallic plates. From the transmission loss, as a function of angle of incidence and the geometry of the sample, the permittivity and loss tangent are deduced.

This is an illustration of quasi-optical methods in the microwave band. With the growing interest in the upper millimetric band, such methods are likely to be applied to the design of microwave components and, perhaps, of complete microwave systems [43]. General mechanical design and manufacturing techniques have been reviewed by Harvey [44], [45].

Solid-State Devices

The fields of interest here are transistors, diodes in general and tunnel diodes in particular, parametric amplifiers and masers.

Transistors are really beginning to reach microwave frequencies [46] but it is doubtful whether transistors of the types available to date will be a commercial proposition. A real success of transistors at microwave frequencies hinges on the development of new types.

It is probably true to say that none of the semiconductor devices has undergone such a rapid development and wide acceptance as has the tunnel diode [47]; yet, for microwave applications the prospects for the tunnel diode are not too bright. Currently available tunnel diodes are capable of delivering only tens of microwatts of power at the lower microwave frequencies, and although research is taking place in a number of establishments, there does not seem to be much chance of inducing a tunnel diode to deliver more than about one milliwatt of power at the lower microwave frequencies or tens of microwatts at the upper microwave frequencies. Clearly, junctions with much higher current densities are needed. This suggests the use of high

mobility compounds such as gallium arsenide, indium antimonide, etc. Using wafer construction, a number of gallium arsenide diodes have been produced [48]. Although the technology of such diodes is more involved than that of the germanium diodes, these diodes are preferred for microwave applications.

Microwave parametric amplifiers and masers form a very new field. Parametric amplifiers entered the microwave field in 1957 (in the United States) through the work of Suhl and Hines but its rapid progress was made possible as a result of Uhlir's work on parametric diodes. Parallel with the diode work, ferromagnetic amplifiers and parametric beam tubes were developed. The age of maser devices is about the same as that of parametric amplifiers, but the chief impetus to the maser work was received from Bloembergen's (1958, United States) proposal of a three-level solid-state maser.

The interest in parametric amplifiers and masers quickly spread to other countries, and in Great Britain there are a number of establishments working in this field. Parametric amplifiers were discussed in a number of review articles (*e.g.*, [49], [50]). Clarricoats analyzed the traveling-wave ferromagnetic amplifier [51] and Thompson reported on an investigation of spin wave excitation in YIG [52]. Various other aspects of parametric amplification were also studied, such as gain, bandwidth and noise figure [53], circuit aspects [54], electron beam amplifiers [55] and measurements on broad-band amplifiers [56]. Cullen was one of the first to analyze the traveling-wave parametric amplifier [57]. He was able to take into account a number of factors influencing the performance of amplifiers, such as ohmic losses and saturation phenomenon [58], and deduced the consequential maximum gain.

The commercial success of parametric devices at upper microwave frequencies—apart from, possibly, parametric beam devices—depends largely on successful development of parametric diodes with cutoff frequencies in the region of 500 Gc. Thus, for example, efficient harmonic generation in the millimetric band is not possible with presently available diodes because of the proximity of cutoff. On the other hand, it is in the upper millimetric band that microwave tubes are also inefficient and, therefore, diodes as harmonic sources might yet succeed.

A different approach to harmonic generation has been reported by Froome. He obtained good harmonic yield using the nonlinear voltage current and relationship of a mercury arc [59]. Subsequently a somewhat different structure was investigated. Using tungsten anode and calcium cathode in argon atmosphere, Froome was successful in obtaining frequencies in the region of 300 Gc (most recent reports state that nearly 600 Gc were produced) from a source of 35 Gc [60]. The efficiency at present is low (it still compares favorably with any other source available at present), but further work is proceeding.

One of the many reasons why interest in parametric amplifiers is so great and is sustained is because of the very low noise figure realizable with such devices. In this respect, however, masers (despite their greater complexity) have an overwhelming advantage; yet the choice depends on the particular application envisaged [61].

General review articles concerning masers have appeared (*e.g.*, [62], [63]). The activity can be divided into 1) investigation of materials and modes of operation, 2) two level devices, 3) three level devices and applications, 4) ammonia maser oscillators, 5) traveling wave masers, 6) optical and infrared masers.

Although from a scientific point of view masers are a great success, any great commercial success is intimately related to the development of more suitable materials—materials with smaller losses and more favorable relaxation times, materials for operation at elevated temperatures (say 60°K), materials suitable for operation in millimetric band, etc.

Solid-state masers using materials with suitably diluted paramagnetic ions are of chief interest. Of these materials, those containing Cr ions dissolved in Al_2O_3, $K_3Fe(CN)_6$, TiO_2, etc., are preferred. A number of investigators have carried out studies of relaxation phenomena (*e.g.* [64], [65]), while rutile was investigated by Thorp for operation in the millimetric band [66]. Two-level masers are also of interest in the study of properties of paramagnetic ions in a host crystal lattice [67].

A number of engineered versions of masers, mainly for operation at lower microwave frequencies, have been described [68]–[70]. Ditchfield reported on an *X*-band maser operating at 77°K; this is a significant step forward, but, of course, at the expense of efficiency [71]. He also reported one study of variation of relaxation times with temperature.

Cavity masers are substantially (but not necessarily so) narrow-band amplifiers of negative resistance type. For many applications this is a limitation. For this reason there is a wide interest in traveling-wave masers because such amplifiers are reputed to be free from such disadvantages [62], [72]. The preferred construction employs a comb slow-wave structure with ruby as the active material.

The power output of ammonia masers is rather small and the bandwidth very narrow. For this reason such devices are not likely to be used as amplifiers. As an oscillator, the ammonia maser is characterized by a high-frequency stability. The prospect of ammonic maser oscillator as a frequency standard is, therefore, good and at least in one research establishment work is proceeding in that direction [73]. The ultimate frequency stability may probably be of the order of a few parts in 10^{11}.

There are many more fields of interest to those engaged on microwaves: experiments with plasma, cyclotron resonance, etc. The published information is

scarce however and largely confined to review articles. Infrared and optical masers, however, ought to be mentioned [74] because of the large impact such devices are likely to have on our future developments.

Utilization

Apart from a number of industrial applications for a variety of uses [3] and applications in defense projects, microwave links remain the chief consumer of microwave equipment.

There is, however, a widespread interest in communication with and *via* satellites; this is evidenced by the large number of lectures and public talks on the subject. The chief effort in this direction seems to be made in the United States; understandably so, because means for putting satellites in space are not, at present, at the disposal of European countries, except, of course, Russia.

In Britain the activity in satellite communication is largely confined to considerations of electronic equipment, investigations of upper atmopshere and theoretical assessments of future prospects; the published information, however, is very scarce.

There is active interest in long-distance communication by waveguide. In fact it transpires that Great Britain, Japan, Russia and the United States are the four countries where research in this field is taking place on a large scale. But, there is interest in a number of other countries, notably France and Germany.

Progress in the field of long-distance communication by waveguide can be gathered from an article by Karbowiak and Solymar [75]. It appears that for efficient operation a waveguide communication system must utilize frequencies corresponding to the upper millimetric band. The authors also point out that the use of special waveguides—either of the dielectric coated variety or a helical construction—is, from a practical point of view, a necessity. It is also stressed that modulation methods such as PCM, where signals can be regenerated at repeaters spaced about 20 miles is also a necessity, but that PCM is not the only possible choice and, furthermore, for short-haul applications a variety of modulation methods could be used with success.

An incidental application of helical waveguides to the construction of high-Q wavemeters has been suggested by Barlow, *et al.* [76]. In this way the authors propose to overcome harmful interference from other modes and thereby improve the wavemeter performance.

Future Trends and Conclusions

The field of microwaves is now so immense that one approaches the task of a critical review with a certain amount of diffidence. It is difficult, indeed impossible, to read and listen to everything that concerns microwaves and yet give a critical account of the work done even though the work concerns a period of time as short as one year. A great deal of preliminary selection has been carried out and it is therefore possible that someone's important contribution has been overlooked.[5] The material has been selected, and the space devoted to the review adjusted, in the following order: new or controversial matters, work of fundamental character or one that is likely to have a significant impact on future developments, miscellaneous fascinating discoveries and modifications. Fields which have not been included in the review (with some exceptions) are: antennas, propagation and microwave tubes. Whereas most of the space was devoted to the theory of microwaves and waveguides, various component measurements and solid-state devices occupy the remainder of the review. Utilization has also been briefly summarized.

In assessing the future trends, it is important to observe that most of the progress achieved in the microwave field is a direct result of research work. Thus future activity and trends will be, in a way, a reflection of the attitude towards research prevailing throughout the country; the more we are research-minded the greater are likely to be the fruits. It is the attitude of mind that determines the future. Admittedly, research is an expensive business, but some years ago the attitude towards research was such as to justify the saying that research is yet another form of hobby.

While not disparaging the spirit behind such an attitude, the reviewer welcomes the change which has taken place during recent years. For it is only by recognizing that research is an essential part of industrial, and indeed of national activity, that a rapid progress in new fields can possibly be made.

Which fields should be exploited first in order that the harvest be worthwhile? This is the big question asked over and over again by many. The simple answer is that in the majority of cases there can be no answer on the available evidence. In fact, it is wrong to ask such a question; for, what is worthwhile?

It is evident from the large volume of published articles and lectures on research management that there is a growing concern (see, for example, [77]) about most aspects of research. It would appear, however, that the surest way to success in research, if that is the aim, is faith. Faith of one member of a research team in another, faith in our profession, and faith in our aims. There is, really, no other recipe. While nowadays, a lot of fine work is still being done by individual people, research is yet another form of social intercourse, but of a very specialized nature.

It is then the research scientists and engineers themselves that have the right qualifications, training and education to guide and influence the future of research.

[5] In such cases the reviewer would be grateful if his attention were drawn to the fact, so that the contribution could be taken into account at a later date.

Yet, so few of them are in sufficiently senior positions to undertake this duty. In this connection Lord Wolton's words spoken in the House of Lords debate on November 21, 1956, may be recalled without comment:

> If industry wants to get first-class scientists and engineers into its service, it has to make perfectly clear to the scientist what can happen to him in the future. The scientist does not want to stay in the laboratory all his days, and just to give advice. He wants to get into a position in which, equally with the finance director or sales director, he will be able in the board room to influence the policy of the business.

Having now established that any future forecast of research progress is next to impossible, we can now summarize the future trends in the microwave field as follows.

There is an indication that in the next few years the fundamental standards of physics will be reviewed; microwaves will play an important part in it.

There is already a greater understanding of theory of electromagnetic fields, and an even deeper understanding of guided waves in waveguides of arbitrary structure should follow. More and more powerful mathematical techniques (such as the Wiener-Hopf technique) are being applied to the solution of waveguide problems. We should, therefore, witness a solution to many outstanding problems.

Microwave techniques are also developing, but there is a general tendency towards higher and higher frequencies. New developments in the upper millimetric band are anticipated and some of the engineering problems connected with the exploitation of the submillimetric spectrum should be overcome.

Efficient generation of coherent radiation in the millimetric and submillimetric spectra is an outstanding problem whose solution will probably lie in the invention of suitable solid-state devices. We may even witness the closing of the present gap between submillimetric and infrared spectra.

In utilization, the application of microwaves to communication systems will dominate our activity (excepting defense projects). In this connection, satellite communication seems to have captured the imagination of many, chiefly because satellites will be launched and rocket research will proceed irrespectively of any other uses, and also because the problems are fascinating and there is a great deal to learn. The commerical success, however, is at present doubtful, but in the infrared or the visible spectrum the chances are perhaps better.

Long distance communication by waveguide will become a reality in a not-too-distant future, particularly since the engineering in the upper millimetric band has reached, now, a high degree of development.

While on the subject of communication systems, it should be observed that although telephone traffic continues to increase, other uses (such as television, data transmission, etc.) which the communication systems serve, increase even faster. It may be that in the years to come telephone traffic will form but an insignificantly small proportion of the activity in communication systems and we will have established even better ways, not yet conceived, of exchanging our ideas.

Acknowledgment

The reviewer gratefully acknowledges the helpful assistance received from many of his colleagues in the preparation of this review. Acknowledgment is also due to Standard Telecommunication Laboratories, Ltd., for help in the preparation of the manuscript and permission to publish it.

References

[1] A. F. Harvey, "Microwave tubes—an introductory review with bibliography," *Proc. IEE*, vol. 107-C, pp. 29–59; March, 1960.
[2] P. C. M. de Belatini, "Inadequacy of scatter mechanisms in tropospheric radio propagation," *Nature*, pp. 1558–1559; November, 1959.
[3] A. F. Harvey, "Industrial, biological and medical aspects of microwave radiation," *Proc. IEE*, vol. 107-B, pp. 557–566; November, 1960.
[4] P. Vigoureux, "Electrical units and standards," *Proc. IEE*, vol. 107-B, pp. 235–240; May, 1960. (A review of progress.)
[5] L. Essen, J. V. L. Parry, and J. McA. Steele, "Frequency variation of quartz oscillators and the earth's rotation in terms of the N.P.L. caesium standard," *Proc. IEE*, vol. 107-B, pp. 229–234; May, 1960.
[6] J. Holloway, *et al.*, "Comparison and evaluation of ceasium atomic beam standard," Proc. IRE, vol. 47, pp. 1730–1736; October, 1959.
[7] P. Vigoureux, "Development of the formulae of electromagnetism in M.K.S. systems," *Proc. IEE*, vol. 107-B, pp. 331–336; July, 1960.
[8] D.S.I.R., National Physical Lab., "The Refractive Index of Air for Radio Waves and Microwaves," London, Eng.; 1960.
[9] A. E. Karbowiak, "Some comments on the classification of waveguide modes," *Proc. IEE*, vol. 107-B, pp. 85–93; March, 1960.
[10] ——, "Radiation and guided waves," IRE Trans. on Antennas and Propagation, vol. AP-7, pp. 191–200; December, 1959. Special supplement.
[11] H. E. M. Barlow, "A physical classification of electromagnetic waves," *Proc. IEE*, vol. 107-B, p. 552 November, 1960.
[12] R. A. Waldron, "Some remarks on waveguide modes," *Proc. IEE*, vol. 108-B, pp. 236–237; March, 1961.
[13] P. J. B. Clarricoats, "Propagation Along Unbounded and Bounded Dielectric Rods," IEE Monograph No. 409E, pt. 1, No. 410E, pt. 2; October, 1960.
[14] E. F. F. Gillespie, "Power flow and negative wave impedance in the dielectric-rod waveguide," *Proc. IEE*, vol. 107-C, pp. 198–201; September, 1960.
[15] W. Hirsch, "The surface wave aerial," *Proc. IEE*, vol. 107-C, pp. 202–212; September, 1960.
[16] J. R. Wait and A. M. Conda, "Resonance excitation of a corrugated-cylinder antenna," *Proc. IEE*, vol. 107-C, pp. 362–366; September, 1960.
[17] C. M. Angulo and W. S. C. Chang, "The Launching of Surface Waves by a Magnetic Line Source," IEE Monograph No. 411E; October, 1960.
[18] E. W. Williams, "Propagation of electromagnetic surface waves along wedge surfaces," *Quart. J. Mech. Appl. Math.*, vol. 13, pp. 278–284; August, 1960.
[19] L. Solymar, "A note on the optimum design of non-uniform transmission lines," *Proc. IEE*, vol. 107-C, pp. 100–104; March, 1960.
[20] D. Wray and R. A. Hastie, "Waveguide bend," *Electronic Technol.*, vol. 37, pp. 76–83; February, 1960.
[21] K. Schnetzler, "The rotation of the plane of polarization of a twisted square waveguide," *Frequenz*, vol. 14, pp. 123–126; April, 1960.
[22] H. E. M. Barlow, "Non-reflecting waveguide tapers," *Proc. IEE*, vol. 107-B, pp. 515–521; November, 1960.

[23] L. Solymar and C. C. Eaglesfield, "Design of mode transducers," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-8, pp. 61–65; January, 1960.
[24] C. C. Eaglesfield, X. Klinger, and L. Solymar, "A new H_{10}-to-H_{20} mode transducer," *Proc. IEE*, vol. 107-B, pp. 512–514; November, 1960.
[25] H. Larsen, "Transmission properties of H_{01} waveguides with statistically distributed irregularities," *Frequeur*, vol. 14, pp. 135–142; April, 1960.
[26] K. Foster, and A. C. Brown, "The application of printed circuit techniques to microwave systems," *Brit. Commun. and Electronics*, vol. 7, pp. 584–589; August, 1960.
[27] J. C. Parr, "Printed circuit waveguides and their application to microwave aerials," *Brit. Commun. and Electronics*, vol. 8, pp. 20–24; January, 1961.
[28] L. Lewin, "Radiation from discontinuities in strip line," *Proc. IEE*, vol. 107-C, pp. 163–170; September, 1960.
[29] J. K. Sinha, "A method for the evaluation of equivalent circuit parameters of an assymetric waveguide junction," *Proc. IEE*, vol. 107-C, pp. 324–329; May, 1960.
[30] L. Lewin, "Some observations on waveguide coupling through medium sized slots," *Proc. IEE*, vol. 107-C, pp. 171–178; September, 1960.
[31] J. R. C. Twisleton, "Some properties of travelling-wave resonance," *Proc. IEE*, vol. 107-B, pp. 108–118; March, 1960.
[32] R. A. Waldron, "Features of cylindrical waveguides containing gyromagnetic media," *J. Brit. IRE*, vol. 20, pp. 695–705; September, 1960.
[33] L. Lewin, "A ferrite boundary value problem in a rectangular waveguide," *Proc. IEE*, vol. 106-B, pp. 559–563; November, 1959.
[34] W. Hauser and K. Brown, "The size effect of ferrite spheres," *Quart. J. Mech. and Appl. Math.*, vol. 13, pp. 257–271; August, 1960.
[35] R. A. Waldron, "Perturbation theory of resonant cavities," *Proc. IEE*, vol. 107-C, pp. 272–274; September, 1960.
[36] I. A. Harris, "A coaxial film bolometer for the measurement of power in the UHF band," *Proc. IEE*, vol. 107-B, pp. 67–72; January, 1960.
[37] I. Lemco, and B. Rogal, "Resistive-film milliwattmeters for the frequency bands 8.2–12.4 Gc/s, 12.4–18 Gc/s and 26.5–40 Gc/s," *Proc. IEE*, vol. 107(B), pp. 427–30; September, 1960.
[38] R. V. Harrowell, "Elliptic waveguide windows—design of resonant types," *Electronic Technol.*, vol. 37, pp. 163–166; April, 1960.
[39] M. Y. El-Ibiary, "Q of resonant cavities—measurement by phase-shift method," *Electronic Technol.*, vol. 37, pp. 284–286; July, 1960.
[40] J. M. Free and G. B. Walker, "Determination of the dielectric properties of low-loss ceramics at Q-band frequencies," *Proc. IEE*, vol. 107-B, pp. 354–356; July, 1960.
[41] J. K. Sinha and J. Brown, "A new cavity-resonator method of measuring permittivity," *Proc. IEE*, vol. 107-B, pp. 522–530; November, 1960.
[42] J. I. Caicoya, "Measurement of dielectric constants by the high-order-mode interferometer," *Brit. Commun. and Electronics*, vol. 7, pp. 32–34; January, 1960.
[43] L. Lewin, "Some comments on quasi-optical methods at millimetric wavelengths," *Proc. IEE*, vol. 107-B, pp. 91–00; March, 1960.
[44] A. F. Harvey, "Mechanical design and manufacture of microwave structure," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, MTT-7, pp. 402–422; October, 1959.
[45] L. E. Hall and B. H. Meggs, "Electroformed copper-waveguides," *Metal Ind. London*, pp. 435–438: November, 1960.
[46] J. Bickley, "Measurement of transistor characteristic frequencies in the 20–1,000 Mc/s range," *Proc. IEE*, vol. 107-B, pp. 301–304; May, 1960.
[47] R. W. A. Scarr, "The tunnel diode—A significant new semiconductor device," *Brit. Commun. and Electronics*, vol. 7, pp. 254–257; April, 1960.
[48] K. G. Hambleton, J. J. Low, and Sherwell, R. J., "Gallium Arsenide Tunnel Diodes," *Nature*, vol. 185, pp. 676–677; March, 1960.
[49] C. R. Russell, "The parametric amplifier," part I, *Brit. Commun. and Electronics*, vol. 7, pp. 94–98; February, 1960. Part II, vol. 7, pp. 190–194; March, 1960.
[50] K. W. H. Stevens, "Parametric amplification," *J. Sci. Instr.*, vol. 37, pp. 1–5; January, 1960.
[51] P. J. B. Clarricoats, "The gain of travelling-wave ferromagnetic amplifiers," *Proc. IEE*, vol. 106-C, pp. 165–173; September, 1959.
[52] A. F. H. Thompson, "Spin Wave Excitation," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April 6–8, 1960.
[53] J. D. Pearson and J. E. Hallett, "Comparison of gain, bandwidth and noise figure of variable reactance amplifiers and convertors," *Proc. IEE*, vol. 107-B, pp. 305–310; May, 1960.
[54] E. L. Neufeld and C. S. Aitchison, "Circuit Design," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April 6–8, 1960.
[55] G. O. Chalk, "Electron-beam parametric amplifier," *Proc. IEE*, vol. 108-B, pp. 125–132; January, 1961.
[56] A. B. McNaughton, "Design an Initial Measurements of a Broad-Band Variable Capacitance Amplifier," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April 6–8, 1960.
[57] A. L. Cullen, "Theory of travelling-wave parametric amplifier," *Proc. IEE*, vol. 107-B, pp. 101–107; March, 1960.
[58] A. Jurkus and P. N. Robson, "Saturation effects in a travelling-wave parametric amplifier," *Proc. IEE*, vol. 107-B, pp. 119–122; March, 1960.
[59] K. D. Froome, "A new harmonic generator," *Nature*, vol. 184; September, 1959.
[60] ——, "Millimetric waves from Mercury Arc harmonic generators," *Nature*, vol. 186, p. 959; June, 1960, and vol. 188, pp. 43–44; October, 1960.
[61] D. C. Laine, "Masers or parametric amplifiers," *Electronic Technol.*, vol. 37, pp. 174–185; May, 1960.
[62] P. N. Butcher, "An introduction to the theory of solid state masers with particular reference to the travelling wave maser," *Proc. IEE*, vol. 107-B, pp. 341–351; July, 1960.
[63] S. A. Ahern, "Solid-state maser amplifier," *Electronic Technol.*, vol. 37, pp. 59–63; February, 1960.
[64] D. H. Paxman, "Spin-Lattice Relaxation Phenomena in Dilute Crystals of $K_3(CN)_6$," Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April 6–8, 1960.
[65] G. S. Waters, "An Experimental Study of Long Spin Lattice Relaxation Times in Doped Silicon Irradiated Calcite," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April 6–8, 1960.
[66] J. S. Thorp, "8-mm Measurements on Maser Materials," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April 6–8, 1960.
[67] D. Bijl, "Experiments on Two-Level Masers," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April 6–8, 1960.
[68] H. W. Duckworth, "The Performance of a 10-CM Radar Incorporating a Maser," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April, 6–8, 1960.
[69] S. A. Ahern, "Packaged S-Band Cavity Maser," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April 6–8, 1960.
[70] W. L. Birch, "Preliminary Results on S-Band Parametric Amplifiers," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April 6–8, 1960.
[71] C. R. Ditchfield, "The Design of an X-Band Maser to Operate at 77°K," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, England, April 6–8, 1960.
[72] J. C. Walling, "Travelling Wave Masers," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April 6–8, 1960.
[73] A. M. J. Mitchell, K. G. Roots, and G. Phillips, "Ammonia maser oscillator. Construction and performance," *Electronic Technol.*, vol. 37, pp. 136–143; April, 1960.
[74] J. H. Sanders, "Optical Masers," presented at Nottingham Conference on "Solid State Microwave Amplifiers," Nottingham, Eng.; April 6–8, 1960.
[75] A. E. Karbowiak and L. Solymar, "Characteristics of waveguides for long-distance transmission," *J. Research Natl. Bur. Standards*, vol. 65, pp. 75–88; January-February, 1961.
[76] H. E. M. Barlow, H. G. Effery, and R. H. Hargrave, "The use of a wire-wound helix to form a circular H_{01} wavemeter cavity," *Proc. IEE*, vol. 107-B, pp. 66; January, 1960.
[77] B. E. Noltink, "The Human Element in Research Management," Elsevier, Amsterdam, Holland; 1959.

A New Type of Circular Polarizer Using Crossed Dipoles*

M. F. BOLSTER†, MEMBER, IRE

Summary—A method of obtaining a circularly-polarized wave by use of two orthogonal dipoles driven in parallel by a common transmission line is shown. The lengths of the dipoles are so chosen that the real part of their input admittances are equal and the angle of the input admittances differ by 90°. When these two conditions are met the resulting radiated wave in a normal direction will be circularly polarized.

The method is applicable both to a circularly-polarized radiating antenna and to the problem of producing a circularly-polarized wave of the TE_{11} mode in a round waveguide. For the first case, an analysis and a method of design are shown, and for the second case an experimentally developed example is given. The second case employs monopoles rather than dipoles for convenience in energizing from a coaxial line.

Introduction

IT has been common practice to produce circular polarization using crossed dipoles.[1] Previous methods have used identical resonant (approximately a half wavelength long) dipoles in each plane, and the proper power and phase relationships in the dipoles were obtained by use of an input phasing and matching network so as to drive the two dipoles from a common generator.

The method presented here uses two unequal length dipoles connected in parallel to a common generator and requires no matching network. The required power and phase relationships are obtained by proper choice of the two dipole lengths. For the first case (dipoles in free space), the lengths are calculable; for the second case (monopoles in waveguide), the lengths must be determined experimentally.

Condition for Circular Polarization

In Fig. 1 are shown two orthogonal cylindrical dipoles fed in parallel and a representative equivalent circuit. The mutual interaction between the two dipoles should be negligible as long as their lengths are much greater than their radii and each dipole is placed in the neutral (or zero) potential plane of the other dipole (as is shown on Fig. 1).

The calculated input admittance of a single dipole as a function of its length is shown in Fig. 2 for lengths in the neighborhood of a half wavelength. These values of admittance are second-order calculated values as

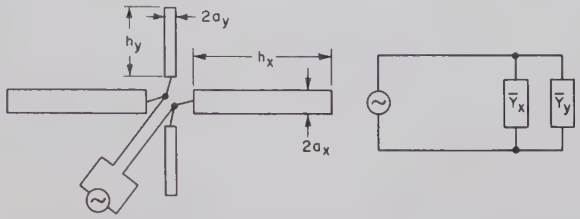


Fig. 1 —Arrangement of dipoles and equivalent circuit.

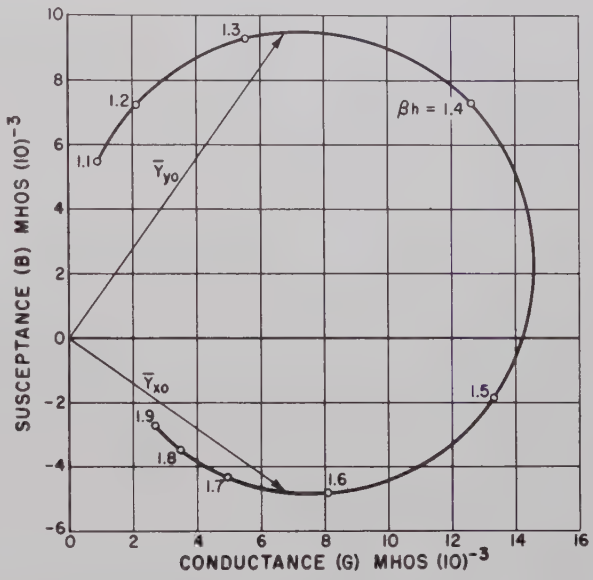


Fig. 2—Dipole input admittance.

given by King[2] for

$$\frac{h_x}{a_x} = \frac{h_y}{a_y} = 75.2.$$

In general, the two dipoles will produce an elliptically-polarized wave. In order to have a circularly-polarized wave, there must be equal power input to each dipole and 90° phase difference between the input currents. This is equivalent to having $G_x = G_y$ and arg $\overline{Y}_x = \arg \overline{Y}_y \pm 90°$ (where $\overline{Y}_x = G_x + JB_x \overline{Y}_y = G_y + JB_y$). This condition can be determined from an accurate plot of dipole admittance by a graphical solution as shown in Fig. 2.

* Received by the PGMTT, April 16, 1961.

† General Engineering Lab., General Electric Co., Schenectady, N. Y.

[1] G. H. Brown, "The turnstile antenna," *Electronics*, vol. 9, pp. 14–17; April, 1936.

[2] R. W. P. King, "The Theory of Linear Antennas," The Harvard University Press, Cambridge, Mass., p. 172; 1956.

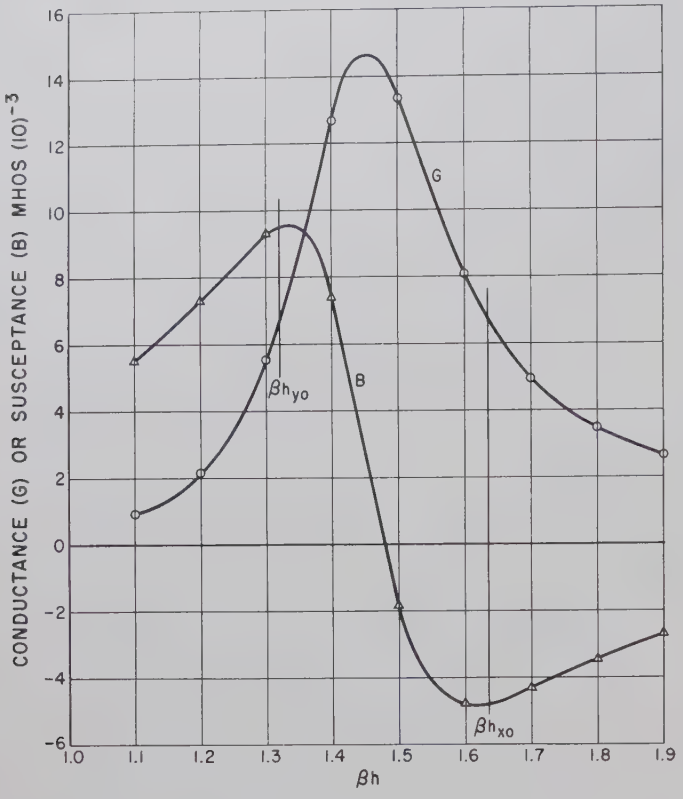


Fig. 3—Real and imaginary parts of dipole admittance.

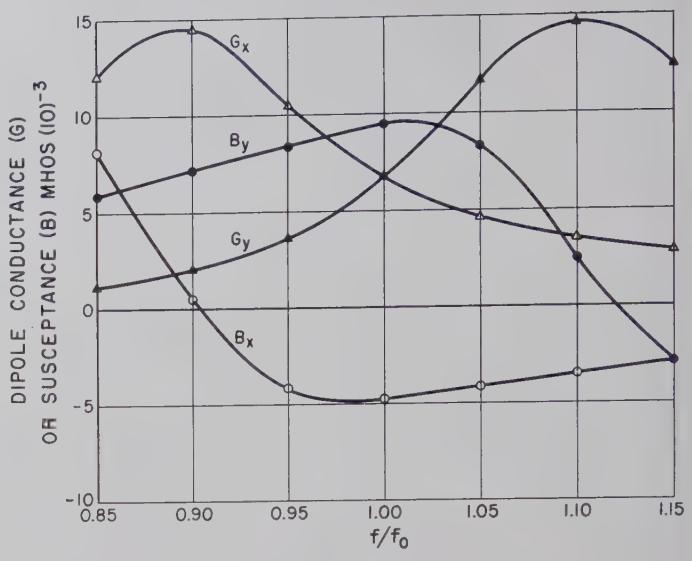


Fig. 4—Input admittance of individual dipoles as a function of frequency.

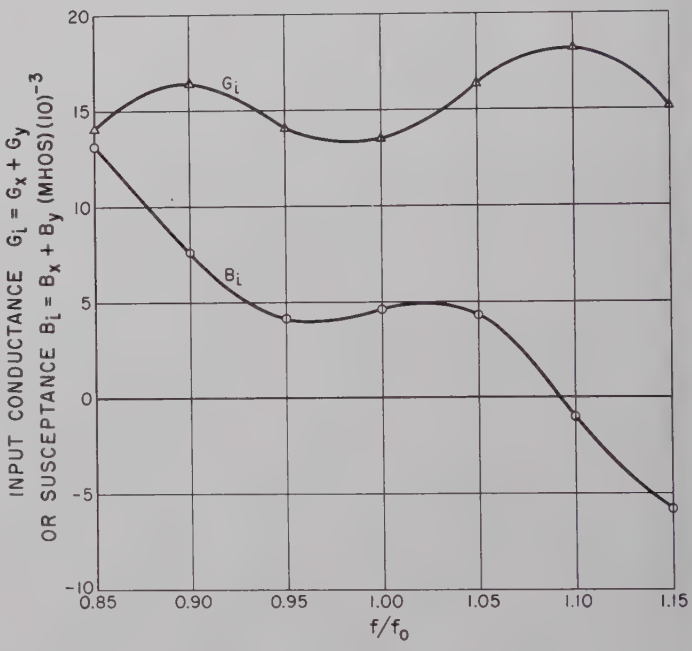


Fig. 5—Input admittance of dipoles connected in parallel.

If the admittance locus were extended to show longer dipole lengths, it would be evident that there are many other combinations of (longer) lengths that would also produce circular polarization.

Variations with Frequency

When the frequency of the driving signal is changed the radiated wave will become elliptically rather than circularly polarized. To determine how the input admittance and ellipticity varies as a function of frequency, it is convenient to replot the real and imaginary parts of the admittance in Fig. 2 as functions of $\beta h = (2\pi/\lambda)h$ as shown in Fig. 3.

From Fig. 3 the two dipole admittances $\overline{Y}_x = G_x + jB_x$ and $\overline{Y}_y = G_y + jB_y$ can be obtained graphically as functions of frequency by simply reading values of $\beta h_x = \beta_0 h_x (f/f_0)$ and $\beta h_y = \beta_0 h_y (f/f_0)$, where f_0 is frequency at which circular polarization occurs. Performing this operation over a frequency range of $f = f_0 \pm 15$ per cent the values of G and B shown in Fig. 4 are obtained.

The total input admittance $\overline{Y}_i = G_i + jB_i = \overline{Y}_x + \overline{Y}_y$ is shown as a function of frequency in Fig. 5.

Polarization Ellipse of Radiated Wave

From the data of Fig. 4 the polarization ellipse of the radiated wave can be calculated. The power in the x or y component waves will be proportional to G_x or G_y, and the amplitude of the component electric fields $\overline{E}_x$ or $\overline{E}_y$ will be proportional to $\sqrt{G}_x$ or $\sqrt{G}_y$. Other nomenclature concerning the polarization ellipse shown in Fig. 6 is defined as follows:

E_x = electric field component in x (dipole) direction
E_y = electric field component in y (dipole) direction
E_x' = electric field component in x' direction
E_y' = electric field component in y' direction
τ = polar angle between xy and $x'y'$ frames
τ_0 = polar angle between dipoles and principal axis of polarization ellipse
$A_x = \sqrt{G}_x$ = amplitude of electric field component in x (dipole) direction
$A_y = \sqrt{G}_y$ = amplitude of electric field component in y (dipole) direction
$R = A_y/A_x$

δ_x = phase of electric field component in x (dipole) direction

δ_y = phase of electric field component in y (dipole) direction.

The field in the x' direction will be

$$E_x' = E_x \cos \tau + E_y \sin \tau, \tag{1}$$

where

$$E_x = A_x \cos (\omega t + \delta_x) \quad \text{and} \quad E_y = A_y \cos (\omega t + \delta_y). \tag{2}$$

This can be written

$$E_x' = A_x' \cos (\omega t + \delta_x'), \tag{3}$$

where

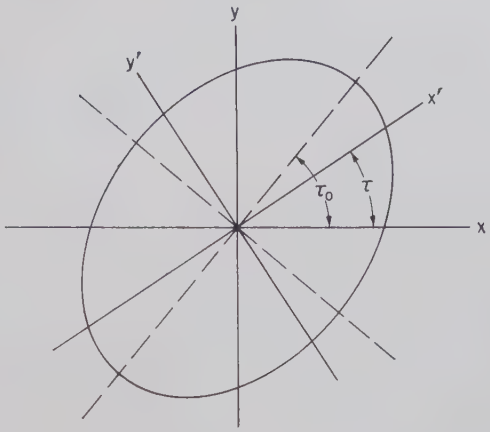


Fig. 6—Polarization ellipse.

$$A_x' = \sqrt{(A_x \cos \delta_x \cos \tau + A_y \cos \delta_y \sin \tau)^2 + (A_x \sin \delta_x \cos \tau + A_y \sin \delta_y \sin \tau)^2}, \tag{4}$$

and

$$\delta_x' = \tan^{-1} \frac{A_x \sin \delta_x \cos \tau + A_y \sin \delta_y \sin \tau}{A_x \cos \delta_x \cos \tau + A_y \cos \delta_y \sin \tau}. \tag{5}$$

The angle τ_0 between the x dipole and the principal axis of the polarization ellipse is obtained by differentiating A_x' with respect to x, setting the result equal to zero, and solving for $\tau = \tau_0$. Carrying out these operations and simplifying the result yields

$$\tau_0 = 1/2 \tan^{-1} \frac{2R \cos \delta}{1 - R^2}, \tag{6}$$

where

$$R = A_y/A_x \quad \text{and} \quad \delta = \delta_y - \delta_x. \tag{7}$$

This is the same formula which Born[3] obtained by a different method.

The axial ratio of the polarization ellipse is defined as

$$R_0' = \frac{A_y'}{A_x'} \Big/_{\tau=\tau_0}, \tag{8}$$

where A_x' is given by (4), and A_y' is obtained in similar manner from

$$E_y' = -E_x \sin \tau + E_y \cos \tau = A_y' \cos (\omega t + \delta_y') \tag{9}$$

$$A_y' = \sqrt{(-A_x \cos \delta_x \sin \tau + A_y \cos \delta_y \cos \tau)^2 + (-A_x \sin \delta_x \sin \tau + A_y \sin \delta_y \cos \tau)^2}. \tag{10}$$

After some simplification, one obtains

$$R_0'^2 = -1 + \frac{1 + R^2}{\cos^2 \tau_0 + R^2 \sin^2 \tau_0 + 2R \cos \tau_0 \sin \tau_0 \cos \delta}. \tag{11}$$

[3] Max Born, "Optik," Verlag Julius Springer, Berlin, Germany, p. 23; 1933.

The axial ratio is obtained from this equation when

$$1/2 \tan^{-1} \frac{2R \cos \delta}{1 - R^2}$$

is substituted for τ_0. This calculation was performed using the data of Fig. 4 and gave the axial ratio (in db) and orientation of the polarization ellipse shown in Fig. 7 over the 30 per cent band.

If the signs of $\cos \delta$ and $1 - R^2$ in (6) are considered, then the angle τ_0 would have to be in the second quadrant for values of $f > f_0$; however, due to the symmetry of the polarization ellipse, plotting τ_0 in the fourth quadrant as shown will cause no error in describing the ellipse.

The shape of the curve for τ_0 shown in Fig. 7 in the region of $f = f_0$ would probably not be duplicated in measurements on an actual antenna. It is unlikely that there would be any cusp at this point. The cusp probably appears in the calculated values because small deviations in the input values of R and δ in the neighborhood of $f = f_0$ result in relatively large changes in the calculated values of τ_0 due to the indeterminate form of (6).

$$\lim_{\substack{R \to 1 \\ \delta \to 90^\circ}} \left[\frac{\cos \delta}{1 - R^2} \right] \to \frac{0}{0} \to 0. \tag{12}$$

The order of the zero of $\cos \delta$ is greater than that of $1 - R^2$ when R and δ are considered as functions of frequency.

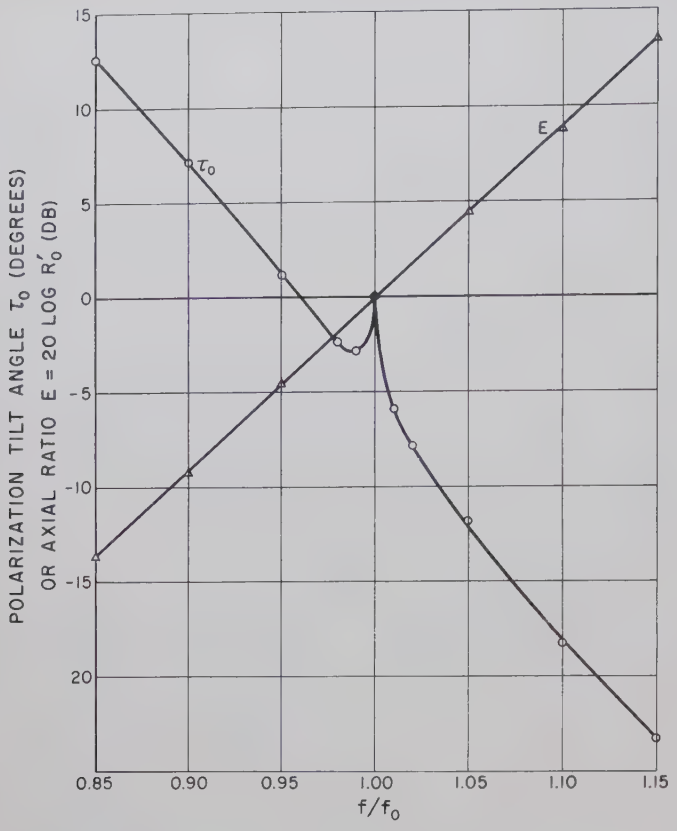


Fig. 7—Axial ratio and tilt angle of polarization ellipse.

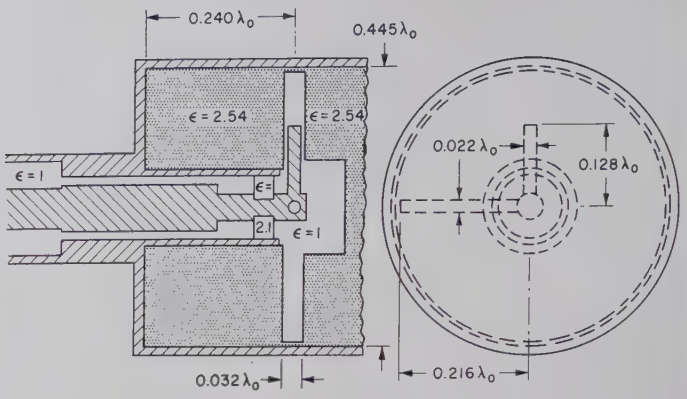


Fig. 8—Circular polarizer in dielectric-filled round waveguide with input matching transformer.

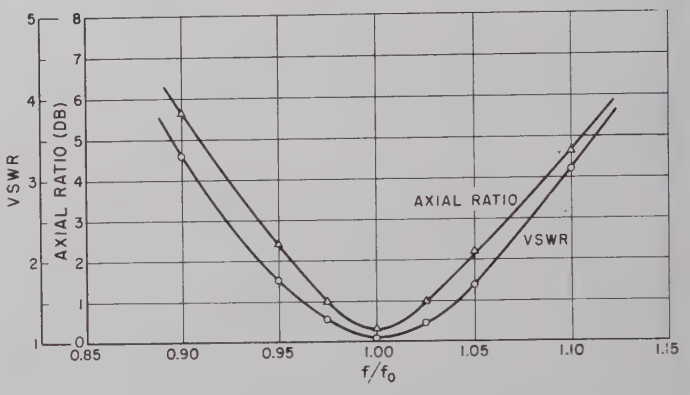


Fig. 9—VSWR and axial ratio of polarizer in dielectric-filled round waveguide with input matching transformer.

Application to Waveguide

The same technique can also be used to produce circular polarization in round waveguide, however, the characteristics are not as readily calculable. Consider a pair of crossed monopoles mounted at 90° to each other and connected together at one end. The two could then be fed in parallel from the center conductor of a coaxial line. If the assembly is next placed inside a round waveguide it is possible to obtain circular polarization in the waveguide by experimental adjustment of the monopole parameters.

An experimentally-adjusted circular polarizer of this type using a dielectric-filled waveguide is shown in Fig. 8. This polarizer was designed to operate over a 10 per cent frequency band from 3000 to 3300 Mc. The monopole elements are arranged to rotate in a circular slot in the dielectric ($\epsilon = 2.54$), and the waveguide diameter (1.666 inch) was chosen so as to propagate only the TE_{11} mode over the desired band.

For the polarizer and coaxial matching transformer shown in Fig. 8, the measured input VSWR and axial ratio of the wave propagated in the waveguide are shown in Fig. 9 for a ± 10 per cent frequency band centered at $f_0 = 3.15$ Gc.

It may be noted that there is approximate agreement in the rate of change of ellipticity between the simple dipoles and the more complex waveguide polarizer of Fig. 8. For a bandwidth of $f_0 \pm 10$ per cent, the dipole ellipticity is about 9 db (Fig. 7), while it is about 5 or 6 db for the waveguide polarizer (Fig. 9).

Conclusions

It is possible to obtain circular polarization by proper choice of the dimensions of the radiating elements themselves, rather than by use of power splitting and phasing networks. The principle is applicable to any type of elements which have input admittance characteristics similar to the dipole of Fig. 2, and it may be accomplished either in free space or inside waveguide.

The free-space type of polarizer would be useful as a circularly-polarized antenna. The fact that no input-phasing network is necessary is a definite advantage. The waveguide type could be used as a waveguide feed for a horn antenna or as a phase shifter by mechanically rotating the monopoles about the axis of the coaxial line. The small moment of inertia would allow rapid rotation with better angular stability and smaller driving torque than previous methods (rotating helix or rotating dielectric plate).

Rounded Corners in Microwave High-Power Filters and Other Components*

SEYMOUR B. COHN†, FELLOW, IRE

Summary—Microwave high-power filters must be operated with internal air pressures of at least one atmosphere, or with a good vacuum. Pressures between these extremes result in reduced power-handling ability. The breakdown processes for both high air pressure and vacuum are discussed, and it is made clear that any sharp corner on which the electric field would concentrate must be rounded if high-power operation is to be achieved. For good results in vacuum operation, the surfaces must be especially smooth and free of contamination, while in high-pressure operation, minor irregularities are less important.

Various high-power filter configurations of importance are described, and the structural corners at which electric-field concentrations occur are pointed out. A number of simplified geometries are then shown that can represent the essential portions of the practical structures with sufficient accuracy for ordinary purposes. Formulas and graphs for these simplified geometries are presented that give the ratio of the maximum electric field strength on the boundary to a uniform reference field strength at a point sufficiently removed from the corner. In some cases, the boundary curve is an approximation to a circular arc, while in other cases a boundary shape is derived such that the electric field strength along the curve is constant. These constant-field-strength boundaries are optimum shapes from the standpoint of power-handling ability.

I. Introduction

THE maximum allowable power flow in high-power filters or other components is limited by ionization breakdown in regions of high electric-field concentration. The critical value of electric field above which breakdown occurs depends upon a number of interdependent factors: 1) the composition and pressure of the air or other gas filling the device; 2) the signal frequency; 3) the size and shape of the region over which the electric field approaches its maximum concentration; 4) the presence of nearby conducting surfaces, their shapes, and their spacings; and 5) the pulse length, shape, and repetition frequency.

The breakdown process as a function of the above parameters has been analyzed by Gould.[1] His work shows that for usual sizes of uniform waveguides and coaxial lines operated near standard-atmosphere air pressure, the ratio of the peak RF electric field strength at breakdown in volts per centimeter to the pressure in atmospheres is approximately independent of factors 2)–5), and is near to the previously accepted value of 29,000 volts/cm/unit pressure in atmospheres.[2] Gould has not considered more complex geometries, but his data for coaxial lines indicate that this figure would hold approximately for rounded-corner radii as small as 1 mm at one atmosphere pressure, with the breakdown field strength being higher for smaller radii or smaller pressure. (For constant error, radius times pressure is a constant.) As the pressure is decreased below one atmosphere, factors 2)–5) begin to have much stronger effects. However, since a high-power filter would usually be operated with at least one atmosphere of pressure, it appears a reasonably good approximation to treat the value of 29,000 volts/cm/atmosphere as a constant. Fortunately, this is a conservative approximation, since the effect of factors 2)–5) is to raise rather than lower the breakdown field strength.

At a pressure of the order of one millimeter of mercury, the breakdown field strength as a function of pressure passes through a minimum. At lower pressures, the breakdown field strength increases rapidly. At pressures lower than about 10^{-6} mm Hg, corresponding to a good vacuum, ionization of the remaining air molecules is no longer important and another mechanism of breakdown becomes limiting. Much less is known about RF breakdown in vacuum than at high pressures. However, current evidence indicates that the RF breakdown mechanism in vacuum is *field emission* of electrons from conducting surfaces. Field emission requires field strengths of the order of megavolts per cm.[3] Such field strengths may develop at minute irregularities, thus initiating breakdown, even when the macroscopic field strength is much lower. Because of this it is extremely important that the surface be smooth and free of contamination.

It has thus been established that the attainment of prespecified values of electric field strength will result in breakdown of high-power microwave filters at high pressures. In vacuo, it is not certain whether a definite breakdown field strength can be prespecified, but nevertheless the breakdown power can certainly be increased by minimizing electric-field concentrations. Thus either in high pressure or in vacuo, any edges at which the electric field may concentrate should be rounded. As an aid in determining the degree of round-

* Received by the PGMTT, March, 7, 1961; revised manuscript received, May 25, 1961. This work was performed at Stanford Res. Inst., Menlo Park, Calif., under the support of the Rome Air Dev. Ctr., under Contract No. AF 30(602)-1998.

† Rantec Corp., Calabasas, Calif.

[1] L. Gould, "Handbook on Breakdown of Air in Waveguide Systems," Microwave Associates, Inc., Burlington, Mass. Rept. on Contract Nobsr 63295; April, 1956.
Also, L. Gould and L. W. Roberts, "Breakdown of air at microwave frequencies," *J. Appl. Phys.*, vol. 27, pp. 1162–1170; April, 1956.

[2] H. A. Wheeler, "Nomogram for Some Limitations on High-Frequency Voltage Breakdown in Air," Wheeler Labs., Inc., Monograph No. 17; May, 1953.

[3] W. S. Boyle, P. Kisliuk, and L. H. Germer, "Electrical breakdown in high vacuum," *J. Appl. Phys.*, vol. 26, pp. 720–725; June, 1955.

ing necessary, a number of rounded-corner geometries important in microwave-filter structures have been analyzed by conformal-transformation methods, and the results are given in this paper. For each geometry considered, the maximum field is related to a reference field at a sufficient distance from the rounded edge that the reference field will be essentially uniform. This information will enable filter designers to relate the maximum field strength occurring in a given filter to the power passing through it. The breakdown value of power may then be determined in the case of high air pressure by setting this maximum field strength equal to approximately 29,000 volts/cm multiplied by the pressure in atmospheres. A safety factor should of course be applied to the calculated breakdown power in order to arrive at a safe power rating for the filter.

Another phenomenon which can, under special circumstances, reduce the power transmitted through an evacuated high-power filter is called *multipactor*.[4] Multipactor is a resonant secondary-emission process in an evacuated region that occurs when an electron under the action of an RF electric field has a time of transit between opposite walls of the region that equals approximately one-half the period of an RF cycle. Limited evidence available to this author indicates that in high-power filters utilizing low-Q elements, multipactor would ordinarily absorb no more than a few watts of power, and hence would not greatly affect performance. If high-Q elements are used however, the loss may be large.

II. Typical High-Power Filter Configurations

Fig. 1 shows simplified sketches of a number of basic high-power filter structures in which electric-field concentration occurs at corners. At all outside corners in the E plane, such as those marked A, the field strength would approach infinity if the radius of the corner were made to approach zero. At all inside sharp corners, such as those marked B, the field strength is zero and is hence no problem. Fig. 1(a) is typical of the corrugated, or varying-impedance filter.[5,6] Fig. 1(b) is a modification of Fig. 1(a), in which the angle of the corner is changed to decrease the field concentration.[7] Figs. 1(c) and 1(d) are leaky wall filters, in which slots in the broad walls of the main waveguide open into

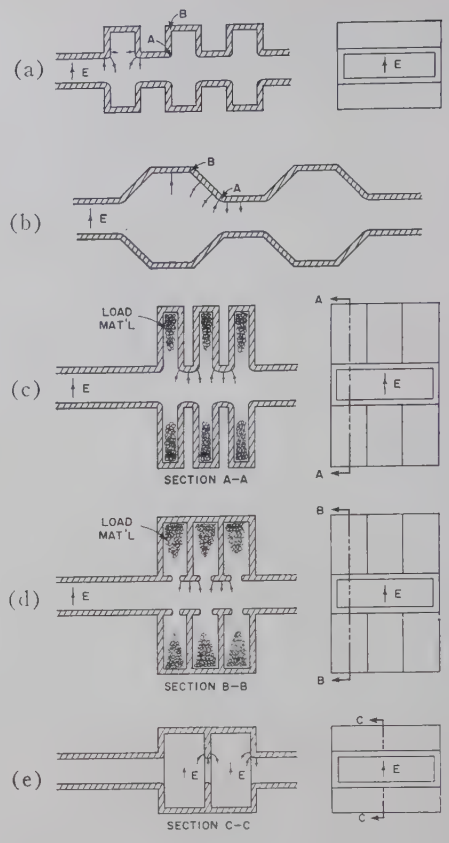


Fig. 1—Typical rounded-corner configurations in high-power filters.

secondary waveguides of higher cutoff frequency.[8–10] Fig. 1(e) is a coupled-cavity filter with rounded edges at the apertures.[8]

Rigorous solutions of the structures in Fig. 1 would be virtually impossible to achieve. However, the solutions for a number of simpler cases shown in Fig. 2 have been obtained that may be applied to the actual filter structures with accuracy sufficient for practical purposes. The two most important assumptions are: 1) the rounded corner geometry may be considered to be composed of infinite cylindrical surfaces, so that the solution for a single two-dimensional cross section is sufficient, and 2) the essential portions of the cross section are small compared to a wavelength, so that the field distribution in those regions may be considered to be very nearly a static field distribution. The first assumption applies exactly to cases like Fig. 1(a) and 1(b), and with fair accuracy to cases like Fig. 1(c)–1(e). The second assumption will usually apply with good accuracy in the pass band of these filters, even though

[4] W. G. Abraham, "Interactions of Electrons and Fields of Cavity Resonators," Ph.D. dissertation, Dept. of Elec. Engrg., Stanford Univ., Stanford, Calif.; 1950.

[5] S. B. Cohn, "Design relations for the wide-band waveguide filter," Proc, IRE, vol. 38, pp. 799–803; July, 1950.

[6] B. M. Schiffman and S. B. Cohn, "Wide-Stop-Band Waveguide Filters," presented at the 1959 PGMTT Natl. Symp., Harvard Univ., Cambridge, Mass.; June 1, 1959.

Also see S. B. Cohn, *et al.*, "Design Criteria for Microwave Filters and Coupling Structures," Stanford Res. Inst., Menlo Park, Calif., Tech. Rept. No. 2, Contract No. DA36-039-SC-74862; June, 1958.

[7] J. H. Vogelman, "High-power microwave filters," IRE Trans. on Microwave Theory and Techniqes, vol. MTT-6, pp. 429–439; October, 1958.

[8] S. B. Cohn, "Design considerations for high-power microwave filters," IRE Trans. on Microwave Theory and Techniques, vol. MTT-7, pp. 149–153; January, 1959.

[9] V. Met, "Absorptive filters for microwave harmonic power," Proc. IRE, vol. 47, pp. 1762–1769; October, 1959.

[10] V. G. Price, R. H. Stone, and V. Met, "Harmonic suppression by leaky-wall waveguide filter," 1959 IRE WESCON Convention Record, pt. 1, pp. 112–118.

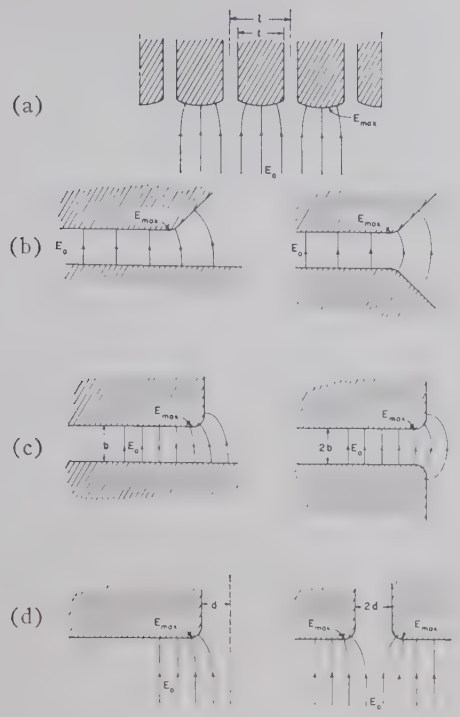


Fig. 2—Two-dimensional geometries considered in this paper.

the over-all waveguide or cavity dimensions are large parts of a wavelength.

The two-dimensional geometries considered in this paper are shown in Fig. 2. The angular designation applied to each corner in Fig. 2 indicates the change in direction of a tangent line as it is moved around the corner. The array of 180-degree rounded corners of Fig. 2(a) has most direct application to the leaky-wall filters of Fig. 1(c) and 1(d). The individual conductors of the array are assumed to be at the same potential. Their center-to-center spacings are assumed to be sufficiently smaller than the spacing to any other conducting surface that the field well below the array of corners may be considered uniform, and equal to a reference value E_0. The 45-degree rounded corner of Fig. 2(b) applies to the varying-impedance filter of Fig. 1(b). By image arguments, the unsymmetrical and symmetrical forms of the 45-degree corner may, of course, be shown to have identical solutions. In this case, the reference field is the uniform field E_0 in the parallel-plane region at a point sufficiently removed from the corner. The 90-degree corner near an electric wall, Fig. 2(c), applies to varying-impedance filter structures when the spacing between the broad walls of the main waveguide is small compared to the width of the slots. The 90-degree rounded corner near a magnetic wall, Fig. 2(d), may be applied to slots in the broad wall of a waveguide when the RF voltage across the slot is relatively small. This would usually be true in the pass band for the leaky-wall filters of Figs. 1(c) and 1(d), but not usually in the pass band of the varying-impedance filter of Fig. 1(a). In that case, when the dimensions are about as shown in the figure, the field in the vicinity of the slot opening may be obtained to a good approximation by superimposing the field distributions of Figs. 2(c) and 2(d).

The coupled-cavity filter of Fig. 1(e) is an example of a more complex structure to which the rounded-corner data of this paper may be applied. The maximum electric field strength will occur in the central longitudinal E plane of the filter. In the absence of the coupling apertures, the field would be greatest at the centers of the cavities, and zero along their vertical sides. In the presence of the coupling apertures, however, there will be electric field concentrations along the top and bottom edges of the apertures that could exceed the field strength at the cavity centers unless the edges are sufficiently rounded. The maximum field strength at the input and output couplings of Fig. 1(e) may be determined from the solution for Fig. 2(c), and that at the central coupling from the solution for Fig. 2(a), with $t/l \ll 1$.

III. Array of 180-Degree Rounded Corners

In the case of an array of 180-degree rounded corners, a method of analysis due to Wheeler[11] has made possible the derivation of the curved-boundary shape upon which the E field is constant (see Part A of Appendix). Figs. 3 and 4 show to scale the resulting boundary shapes for different values of plate thickness t to center-to-center spacing l. It is seen that the radius of curvature varies greatly over each curved portion of the boundary, being relatively large at the center point and decreasing toward the ends. This variation in curvature is necessary in order to maintain constant field strength over the curved boundary. On the straight, vertical portions of the boundary, the field drops off rapidly from the corner. If the curved boundary had any other shape, the field would be nonuniform, attaining in some regions values higher than the constant value occurring on the boundaries of Figs. 3 and 4. Consequently, the boundary shapes in these figures are particularly significant for the design of high-power filters requiring this basic configuration.

The constant field strength E_{max} turns out to be very simply related to the uniform field E_0 well below the array, as follows:

$$\frac{E_{max}}{E_0} = \sqrt{\frac{l}{t}}. \tag{1}$$

The square-root relationship makes E_{max}/E_0 increase slowly as t/l is decreased. For example, if t/l equals 0.5, E_{max}/E_0 will equal 1.414, and if t/l equals 0.1, E_{max}/E_0 will equal 3.16. Points x, y on the curved

[11] H. A. Wheeler, "Conformal Mapping of Rounded Polygons by a Wave-Filter Analog," Wheeler Labs., Inc., Great Neck, N. Y., Rept. No. 667; August 8, 1955.

boundary may be computed by means of the following parametric equations:

$$x = \frac{l-t}{2\pi}\left[p \tan^{-1}(p \tan \phi) - \phi\right] \tag{2}$$

$$y = \frac{l-t}{4\pi} \ln\left[1 + (p^2 - 1)\sin^2 \phi\right], \tag{3}$$

where $p=(1+t/l)/(1-t/l)$, and where ϕ is an independent variable. The co-ordinate system is shown in Fig. 5. As ϕ is increased from $-\pi/2$ to $\pi/2$, the point x, y moves from the left end of the curve to the right end, resulting in a curve that is a symmetrical function of ϕ. The coordinates of the end points are

$$\frac{x_1}{t} = \pm\frac{1}{2}, \qquad \frac{y_1}{t} = \frac{1}{2\pi}\frac{1-t/l}{t/l}\ln\frac{1+t/l}{1-t/l}. \tag{4}$$

The value of y_1/t decreases from 0.318 for $t/l=0$ to 0 for $t/l=1$, which results in the curve shapes being more flattened for the larger t/l values, as may be seen in Fig. 4.

There are two special cases of constant-field-strength boundaries that may be derived from (2)–(4). The first is that of an isolated plate with rounded edge, obtained when $l/t\to\infty$. The curve for this case, which was computed previously by Wheeler,[11] is plotted in Fig. 6. The end points are $x_1=\pm t/2$ and $y_1=0.318t$. This curve may be used in other situations than that of Fig. 6 as an approximation to the ideal curve, if the edge of a plate is removed from any other boundary surface by a distance equal to at least several plate thicknesses.

The other special case is that of an isolated slot with rounded edges where the two sides of the slot are at the same potential. The boundary shape, obtained when $t/l\to 1$, is shown in Fig. 7. It is seen that the curve falls monotonically to the left, but approaches zero slope as $x\to-\infty$. Eq. (1) shows that the field strength on the curved boundary is equal to the uniform field strength far below it. However, because of its infinite extent, the boundary must in practice be altered at some point, in which case a somewhat greater maximum field strength will occur.

Fig. 3—Arrays of rounded 180-degree corners shaped for constant electric field.

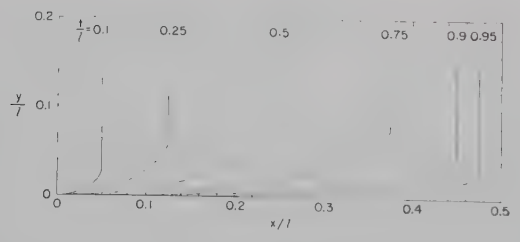


Fig. 4—Shapes of rounded 180-degree corners for various values of t/l.

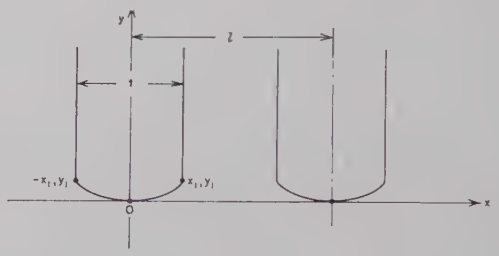


Fig. 5—Coordinate system for array of 180-degree corners.

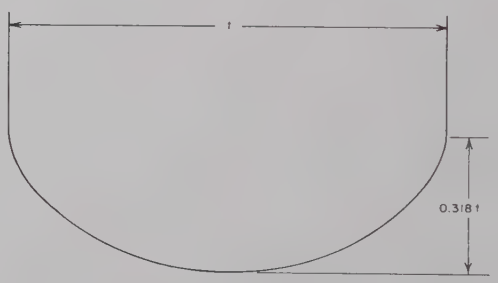


Fig. 6—Isolated rounded 180-degree corner plotted to scale.

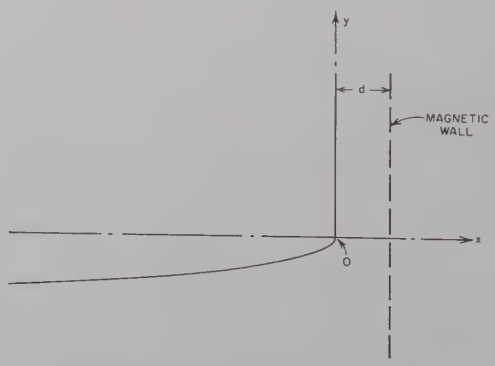


Fig. 7—Limiting case of rounded 180-degree corner for $t/l\to 1$.

IV. Rounded 90-Degree Corner Near Electric Wall

Fig. 2(c) shows a rounded 90-degree corner near an electric wall, and its symmetrical equivalent. Two curved-boundary shapes are of interest: 1) a circular arc, and 2) a shape resulting in constant field strength on the rounded surface.

A. Approximate Circular Boundary

It is not feasible to solve exactly the case of a 90-degree circular-arc corner, but a solution has been obtained for a shape that is a good approximation to a circular arc.[12–14] The actual curved boundary obtained in this case is shown in Fig. 8 for two values of r/b, where r and b are defined in the figure. These boundaries were plotted with the aid of formulas given by Weber.[15] (Note, however, that Weber's plot of one of the boundaries[16] is grossly inaccurate.) The formula for the ratio of E_{max}/E_0 as a function of r/b is also given by Weber, and is plotted in Fig. 9. For example, $E_{max}/E_0 = 1.40$ at $r/b = 0.6$. Thus, in a rectangular waveguide containing such a corner, breakdown will occur at about half the breakdown power of the uniform waveguide itself.

An abrupt step in height of a waveguide, or in diameter of a coaxial line, has an equivalent circuit consisting simply of a shunt capacitive susceptance at reference planes coinciding with the step itself. Graphical data have been published giving the value of this susceptance for sharp-corner rectangular-waveguide and coaxial-line configurations.[17,18] If the corner is rounded, the discontinuity susceptance will be less than that of a sharp corner. A formula for the change in susceptance, ΔB, was previously derived by this author[19] for the approximate circular-arc case. The resulting plot of ΔB versus r/b is given in Fig. 10 for a step in height in rectangular waveguide. The total shunt susceptance of the rounded step is

$$B = B\big|_{r=0} + \Delta B, \tag{5}$$

where $B\big|_{r=0}$ is the susceptance of a sharp step.[16] Note that ΔB is a negative quantity, and hence B is smaller than $B\big|_{r=0}$. Rounding of the corner also results in an

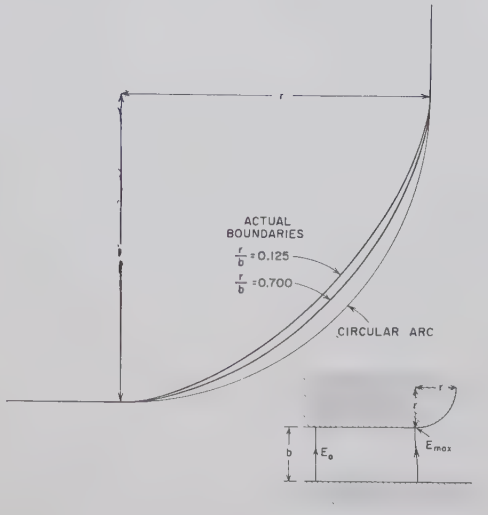


Fig. 8—Actual shape of approximately circular, 90-degree rounded corner near an electric wall.

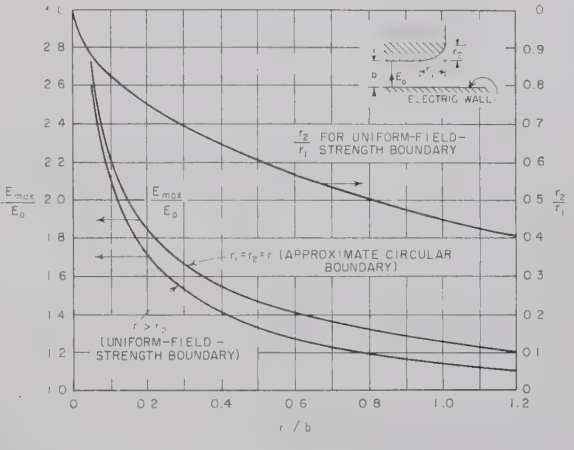


Fig. 9—Plot of E_{max}/E_0 for rounded 90-degree corner near an electric wall; also, r_2/r_1 for a uniform-field-strength boundary.

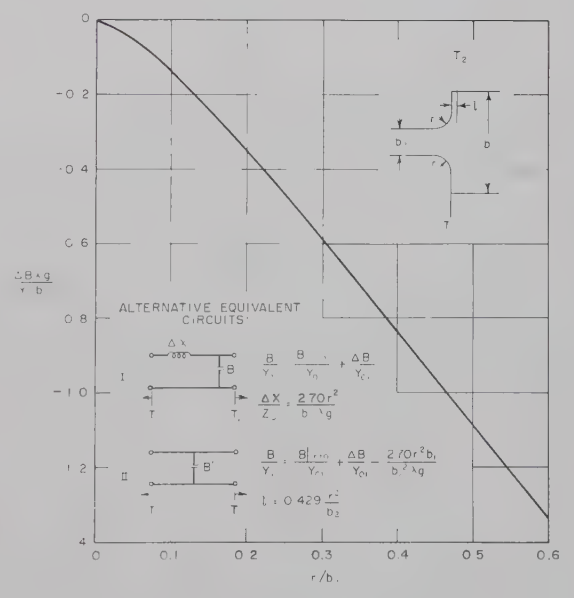


Fig. 10—Capacitive susceptance correction and equivalent circuits for rounded step in waveguide.

[12] L. Dreyfus, "Zur Berechnung von Durchslags und Überschlagsspannung," *Arch. Electrotech.*, vol. 13, p. 131; 1924.
[13] J. D. Cockcroft, "The effect of curved boundaries on the distribution of electrical stress round corners," *J. IEE*, vol. 66, pp. 385–409; 1928.
[14] E. Weber, "Electromagnetic Fields," John Wiley and Sons, Inc., New York, N. Y., vol. I, pp. 373–377; 1950.
[15] *Ibid.*, p. 374.
[16] *Ibid.*, p. 376.
[17] N. Marcuvitz, "Waveguide Handbook," McGraw-Hill Book Co., Inc., New York, N. Y., Rad. Lab. Series, vol. 10, pp. 307–312; 1951.
[18] J. R. Whinnery, H. W. Jamieson, and T. E. Robbins, "Coaxial-line discontinuities," Proc. IRE, vol. 32, pp. 695–709; November, 1944.
[19] S. B. Cohn, unpublished note; March, 1958.

increase in stored magnetic energy. This may be taken account of by adding a series inductive element in the equivalent circuit at the reference plane of the step, or by a shift of one of the reference planes and modification of B. These alternative equivalent circuits are shown in Fig. 10. Values of ΔB taken from the curve in Fig. 10 will be accurate as long as r is sufficiently small compared to the dimension $(b_1-b_2)/2$. Good results should be expected for r as large as one half that dimension. It is also necessary that $\Delta X/Z_{02}$ be small; *e.g.*, less than 0.3 should suffice.

B. Constant-Field-Strength Boundary

A solution in the case of a rounded 90-degree corner near an electric wall has been obtained yielding the boundary shape having uniform field strength on the curved arc.[12,13] Cockroft[13] gives formulas from which this boundary shape may be plotted, and shows in his Fig. 19 one such boundary drawn to scale. The shape of the boundary is very much like that of Fig. 8, but compressed in height. The plot of E_{max}/E_0 versus r_1/b for this case is included in Fig. 9. Also plotted is the ratio of r_1 and r_2, the dimensions that define the horizontal and vertical extent of the curved boundary. As should be expected, the E_{max}/E_0 curve for the uniform-field-strength case falls below that for the approximate circular-arc case. The two cases coalesce as r_1/b approaches zero.

V. Rounded 90-Degree Corner Near Magnetic Wall

A rounded 90-degree corner near a magnetic wall is shown in Fig. 2(d), along with its symmetrical equivalent. As in the case of an adjacent electric wall, Section IV, two curved-boundary shapes are of particular interest: 1) a circular arc, and 2) the shape resulting in constant field strength on the rounded surface.

A. Approximate Circular Boundary

The approximate-circular-arc solution discussed in Section IV-A has been modified to apply to an adjacent magnetic wall (see Part B of Appendix). The boundary shape is exactly the same as in Section IV-A, and hence Fig. 8 applies to the magnetic-wall case as well as to the electric-wall case if $r/d=r/b$. For the adjacent magnetic wall, the ratio of the maximum field strength on the boundary to the distant uniform field strength was found to be

$$\frac{E_{max}}{E_0} = (1+\lambda)\sqrt{1+p}, \tag{6}$$

where λ and p are parameters plotted by Weber[14] in his Fig. 28.8. Eq. (6) is plotted in Fig. 11. The maximum-field-strength point occurs at the end of the curved boundary farthest from the magnetic wall.

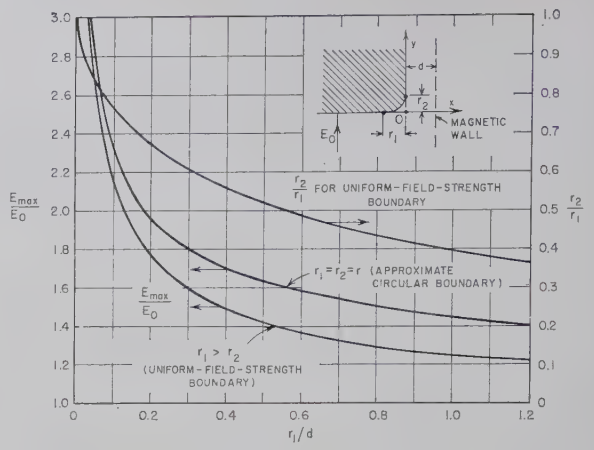


Fig. 11—Plot of E_{max}/E_0 for rounded 90-degree corner near a magnetic wall; also, r_2/r_1 for a uniform-field-strength boundary.

B. Constant-Field-Strength Boundary

The rounded 90-degree corner shape having constant field strength E_{max} on the curved portion of the boundary has been solved for the case of an adjacent magnetic wall [see Part B-2) of Appendix]. The curved-boundary end-point dimensions r_1 and r_2 are given by

$$\frac{r_1}{d} = \frac{1}{\pi}\left[\sqrt{\frac{1}{\lambda^2}-1} - \tan^{-1}\sqrt{\frac{1}{\lambda^2}-1}\right] \tag{7}$$

$$\frac{r_2}{d} = \frac{1}{\pi}\left[\tanh^{-1}\sqrt{1-\lambda^2} - \sqrt{1-\lambda^2}\right], \tag{8}$$

where λ is a parameter with value between 0 and 1. In terms of λ, the ratio of constant field strength E_{max} on the curved boundary to the distant uniform field E_0 is

$$\frac{E_{max}}{E_0} = \sqrt{\frac{1+\lambda}{1-\lambda}}. \tag{9}$$

The resulting curves of E_{max}/E_0 and r_2/r_1 are shown in Fig. 11. Note that the E_{max}/E_0 curve falls below the corresponding curve for the approximate circular-boundary case, as would be expected.

Detailed boundary curves for this case may be plotted from the following equations:

$$\frac{x}{d} = \frac{1}{\pi}\left\{\tan^{-1}\sqrt{u-1} - \sqrt{u-1}\right\} \tag{10}$$

$$\frac{y}{d} = \frac{1}{\pi}\left\{\tanh^{-1}\sqrt{1-\lambda^2 u} - \sqrt{1-\lambda^2 u}\right\}, \tag{11}$$

where λ is the same parameter as in (7), (8), and (9), u is a variable in the interval $1 \le u \le 1/\lambda$, and the coordinate system is as shown in Fig. 11. The boundary shapes yielded by (10) and (11) resemble those in Fig. 8 very closely, if the horizontal and vertical scale factors are altered to make the end points correspond to the appropriate values of r_1/d and r_2/d.

VI. Rounded 45-Degree Corner

The solution for a rounded 45-degree corner adjacent to an electric wall is outlined in Part C of the Appendix. This configuration and its symmetrical equivalent are shown in Fig. 2(b). The resulting curved boundary is neither a circular arc, nor a shape having constant field strength. For $r/b \ll 1$, the curve has the appearance shown in Fig. 12. The formula for this curve is

$$z = -\frac{r}{2^{5/4}}\left\{(t-1)^{5/4} + (t+1)^{5/4}\right\}, \qquad (12)$$

where $z = x + jy$ is a point on the curve with coordinate system as shown in Fig. 12, and t is a real variable in the interval $-1 \leq t \leq 1$. The field strength along the curved portion of the boundary is given, in the case $r/b \ll 1$, by

$$E(z) = \frac{(2+\sqrt{2})^{1/2} E_c}{\left[(1-t)^{1/2} + (1+t)^{1/2} + \sqrt{2}(1-t)^{1/4}(1+t)^{1/4}\right]^{1/2}} \qquad (13)$$

in which E_c is the field strength at the center point of the curve. Because of the assumption that $r/b \ll 1$, $E(z)$ is a symmetrical function of t near the corner. Eq. (13) is plotted in Fig. 13, where it is seen that the field strength on the boundary is almost constant except very near the junction points between curved and straight parts of the boundary. For $r/b \ll 1$, the ratio of field strength at the center of the curve to the uniform field strength in the parallel-plane region is

$$\frac{E_c}{E_0} = \frac{2}{(1+\sqrt{2})^{1/2}}\left(\frac{2}{5\pi}\right)^{1/5}\left(\frac{b}{r}\right)^{1/5} = 0.853\left(\frac{b}{r}\right)^{1/5}. \qquad (14)$$

Clearly, if the transitions from the curved-to-straight portions of the boundary were made slightly more gradual, the sharp rise at the junction points could be greatly reduced, with the maximum field strength on the boundary perhaps 1.1 times the value at the center of the curve.

Because of the one-fifth-root dependence in (14), the increase in E_c/E_0 is very slow relative to b/r. For example, $E_c/E_0 = 1.35$ when $r/b = 0.1$, while for $r/b = 0.01$, $E_c/E_0 = 2.14$. These are much smaller field-strength values than occur with 90-degree rounded corners.

VII. Conclusions

In high-power filters, whether operated with high internal pressure or vacuum, the power-handling ability is limited by electric-field concentrations at corners. By increasing the radius of curvature of these corners, the power rating of the filter may be increased. The formulas and graphs contained in this paper give quantitative data for various practical rounded-corner configurations. With these data, a filter designer can predict the power rating of a given structure or, if a particular rating is specified, he can select the necessary rounded-corner dimensions to meet that requirement.

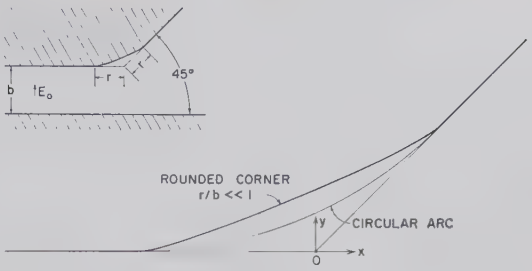


Fig. 12—Shape of rounded 45-degree corner drawn to scale for $r/b \ll 1$.

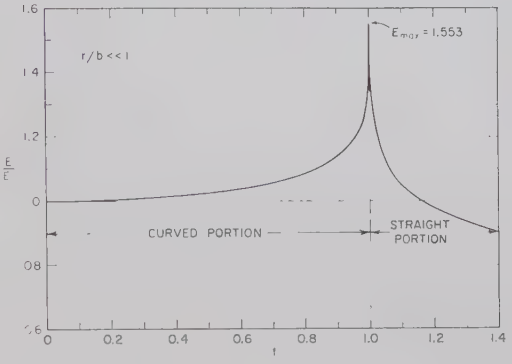


Fig. 13—Electric field strength on boundary of the rounded 45-degree corner shown in Fig. 12 (t is a "length" parameter measured from the center of the corner).

Some of the rounded corners considered in this paper are approximations to circular arcs, while others are shaped so that the electric field strength is uniform on the rounded portion of the boundary. The uniform-field-strength corners may be considered to be optimum, since any other boundary curve starting and ending at the same points would have a greater field strength somewhere on the curve.

Machining with high precision the exact theoretical shapes treated in this paper would not be a straightforward operation. Fortunately, minor deviations in shape will result in only small changes in the value of maximum field strength, as long as the boundary curve is smooth.

Appendix

A. Analysis of Array of 180-Degree Rounded Corners

Following a suggestion of Wheeler,[11] one may show that the periodic z-plane boundary in Fig. 14(a) is transformed into the real axis of the $w = u + jv$ plane in Fig. 14(b) by the following relation:

$$\frac{dz}{dw} = \sqrt{1 - m^2 \tan w} + jm \tan w, \qquad (15)$$

where m is a real constant between 0 and ∞. The single period of the z-plane boundary between $x = 0$ and $x = \pi$

maps into the u axis, also between 0 and π. Inspection of (15) shows that dz/dw is constant in magnitude but variable in angle for w real and in the range $|u| \leq \tan^{-1}(1/m)$. This range of u corresponds to the curved part of the z-plane boundary. The electric field strength in the z plane is proportional to $|dw/dz|$, and hence the field strength is constant along the curve. Further examination of (15) shows that this constant value of field strength is the maximum occurring in the cross section. Integration of (15) and substitution of the appropriate boundary conditions yields (1)—(4).

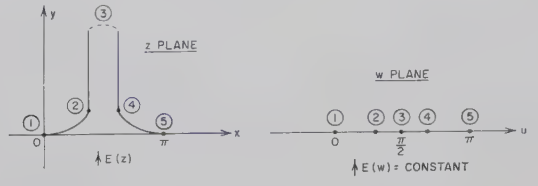


Fig. 14—Transformation for array of 180-degree rounded corners.

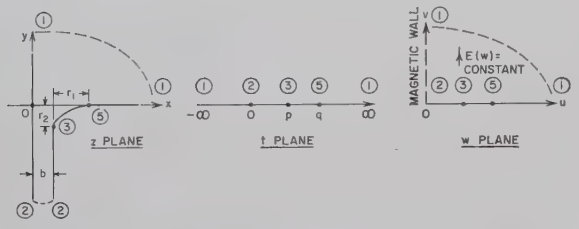


Fig. 15—Transformations for rounded 90-degree corner near magnetic wall.

B. Analysis of Rounded 90-Degree Corner Near Magnetic Wall

1) *Approximate Circular Boundary:* Weber[14] gives the following transformation (with notation changed) relating the z- and t-plane boundaries of Fig. 15:

$$\frac{dz}{dt} = \frac{C_1}{t}\left\{\sqrt{t-p} + \lambda\sqrt{t-q}\right\}, \tag{16}$$

where C_1, p, q, and λ are real, positive constants, with $q > p$. The formula for z is obtained by integration. Relations between p, q, and λ are established such that $r_1 = r_2 = r$. These constants may be computed from his formulas as functions of r/b, or read from his graph.[20] Weber applied this transformation to the case of an electric-wall boundary adjacent to the rounded corner by assuming the real axis of the t plane to be an electric wall with a discontinuity in potential at point 2. However, the same transformation may be used for an adjacent magnetic wall by transforming the t plane into the w plane as shown in Fig. 15. The vertical magnetic wall in the w plane maps into the adjacent magnetic wall of the z plane. Since the electric field in the upper-right-hand corner of the w plane is uniform, the electric field strength in the z plane is proportional to $|dw/dz|$, and was found to be

$$E(z) \propto \frac{1}{\left|\sqrt{1-\frac{p}{t}} + \lambda\sqrt{1-\frac{q}{t}}\right|}. \tag{17}$$

From this, (6) for E_{max}/E_0 was derived.

2) *Constant-Field-Strength Boundary:* Inspection of (17) shows that if λ is set equal to $\sqrt{p/q}$, $E(z)$ will be constant for t real and in the interval $p \leq t \leq q$. When $\lambda = \sqrt{p/q}$ is substituted in (17) and in Weber's formulas, (7)—(11) result.

C. Analysis of Rounded 45-Degree Corner

By application of the Schwartz-Christoffel method, the z-plane boundary in Fig. 16(a) was transformed into the real axis of the t plane by the following relation:

$$\frac{dz}{dt} = \frac{Ct^{1/4}}{t-p}. \tag{18}$$

After integration and substitution of boundary conditions, the following formula for z as a function of t was obtained:

$$z = z(t) = \frac{1+j}{\pi} b\left\{2\sqrt{2}\left(\frac{t}{-p}\right)^{1/4} - \ln\left\{\frac{\left(\frac{t}{-p}\right)^{1/2} + 1 + \sqrt{2}\left(\frac{t}{-p}\right)^{1/4}}{\left(1+\frac{t}{-p}\right)^{1/2}}\right\} - \tan^{-1}\left\{\frac{\sqrt{2}\left(\frac{t}{-p}\right)^{1/4}}{1-\left(\frac{t}{-p}\right)^{1/2}}\right\}\right\}, \tag{19}$$

where p is a positive constant. This transformation yields an abrupt 45-degree corner. To round the corner as in Fig. 16(b), (19) was modified into the following form:

$$z = Az(t') + (1-A)z(t''), \tag{20}$$

where

$$t' = \left(\frac{p+\delta_1}{p}\right)t - \delta_1 \text{ and } t'' = \left(\frac{p-\delta_2}{p}\right)t + \delta_2. \tag{21}$$

The changes in variable of (21) are such that $t' = t'' = t$ at the critical points $t = 1$ and ∞, but not at $t = 0$. At $t' = 0$, $t = p\delta_1/(p+\delta_1)$, while at $t'' = 0$, $t = -p\delta_2/(p-\delta_2)$. As a result of this modification, a gradual change of slope from 0 to 45 degrees occurs with (20) for t in the range $-\delta_2/(1-\delta_2) \leq t \leq \delta_1/(1+\delta_1)$.

[20] Weber, *op. cit.*, p. 375.

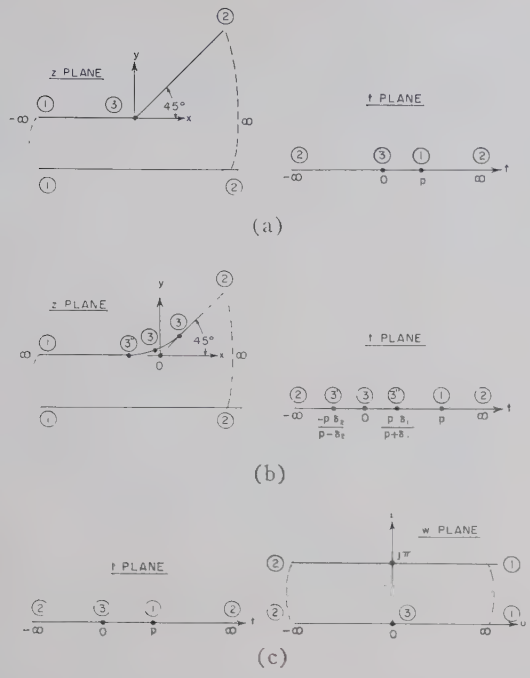


Fig. 16—Transformations for 45-degree corner. (a) Abrupt 45-degree corner. (b) Rounded 45-degree corner. (c) Transformation between t and w planes.

The procedure to follow in determining the contants A, δ_1, and δ_2 for a given value of r/b is to let (20) take on the following values:

$$z = -r \quad \text{when} \quad t' = 0 \quad \text{and} \quad t'' = \frac{p(\delta_1 + \delta_2)}{p + \delta_1} \tag{22}$$

$$z = \frac{1+j}{\sqrt{2}} r \quad \text{when} \quad t' = \frac{p(\delta_1 + \delta_2)}{p - \delta_2} \quad \text{and} \quad t'' = 0. \tag{23}$$

Since these are only two conditions and there are three constants to be determined, one additional condition may be specified. This could be, for example, that the curve should be most nearly symmetrical about its center point.

The case $r/b \ll 1$ is of particular practical interest. In the limit $r/b \to 0$, the constants have values $A = 1/2$ and $\delta_1 = \delta_2 = \delta$, so that δ becomes the only unknown. Furthermore $\delta \ll p$ so that (20) and (21) become

$$z = \tfrac{1}{2} z(t - \delta) + \tfrac{1}{2} z(t + \delta). \tag{24}$$

In the limit $t \to 0$, (19) and (24) reduce to

$$z = -\frac{2b}{5\pi p^{5/4}} \left\{ (t - \delta)^{5/4} + (t + \delta)^{5/4} \right\}. \tag{25}$$

Eq. (25) may be used to calculate the shape of the z-plane boundary in the vicinity of the rounded corner, but it does not, of course, yield the adjacent electric wall.

A discontinuity in potential occurs at point 1 in the z and t planes, so that the electric field in the t plane is nonuniform. The real axis of the t plane may be transformed into the w-plane boundary of Fig. 16(c) containing a uniform electric field by means of this differential relation

$$\frac{dw}{dt} = \frac{1}{p - t}. \tag{26}$$

The electric field within the z-plane boundary near the corner may then be obtained by differentiation of (23) and combination with (26), with $|t| \ll p$,

$$E(z) \propto \left| \frac{dw}{dz} \right| = \frac{2\pi p^{1/4}/b}{\left| (t - \delta)^{1/4} + (t + \delta)^{1/4} \right|}. \tag{27}$$

The uniform field E_0 far from the corner is obtained from (18) and (26) with $t \to p$ and $C = -b/\pi p^{1/4}$

$$E_0 \propto \left| \frac{dw}{dz} \right| = \frac{\pi}{b}. \tag{28}$$

Combination of (27) and (28) gives the following formula for $E(z)/E_0$ valid in the vicinity of the rounded corner for the case $r/b \ll 1$:

$$\frac{E(z)}{E_0} = \frac{2p^{1/4}}{\left| (t - \delta)^{1/4} + (t + \delta)^{1/4} \right|}. \tag{29}$$

The maximum-field-strength points occur at $t = \pm\delta$. At $t = \delta$, (25) yields

$$r = \frac{2b}{5\pi p^{5/4}} (2\delta)^{5/4} \quad \text{and} \quad p^{1/4} = \left(\frac{2b}{5\pi r} \right)^{1/5} (2\delta)^{1/4}. \tag{30}$$

Eqs. (12)–(14) now follow readily from (25), (29), and (30), when δ is set equal to unity.

A Periodic Structure of Cylindrical Posts in a Rectangular Waveguide*

EDWARD E. ALTSHULER†, MEMBER, IRE

Summary—The propagation characteristics of a rectangular waveguide loaded with uniformly spaced cylindrical posts (periodic structure) are investigated at a frequency of 2840 Mc. A qualitative discussion on the expected behavior of the effective guide wavelength of this type of periodic structure is presented, and it is shown that the presence of the posts reduces the guide wavelength of the waveguide. The guide wavelength is then measured as a function of post diameter, post depth, and post spacing; and curves enabling one to design periodic structures which have guide wavelengths in the region of the free space wavelength are presented.

Introduction

THE propagation characteristics of a rectangular waveguide are modified when the guide is loaded with uniformly spaced posts. The purpose of this investigation is to determine how the effective guide wavelength of the periodic structure shown in Fig. 1 varies as a function of post geometry and its depth of penetration.

To understand the behavior of this periodic structure it is first necessary to examine the equivalent circuit of a single element of the structure. A cylindrical post of variable height and diameter in a rectangular waveguide can be represented by the four-terminal network shown in Fig. 2. An approximate theoretical solution for this post has been obtained by Suzuki;[1] however, the final expressions for the reactances are very difficult to evaluate. Experimental results have been obtained by Marcuvitz[2] and although they cannot be scaled directly to the waveguide and post dimensions being investigated, they can be used to describe the qualitative behavior of the post well enough to analyze the periodic structure.

In Fig. 2 jx_b is a relatively small series reactance that increases slightly with post depth and diameter. The main effect of the post is to introduce a shunt reactance jx_a across the waveguide. In Fig. 3 jx_a is plotted as a function of post depth h/b for various post diameters d. As the post penetration is increased, the capacitive reactance of the post approaches zero (when

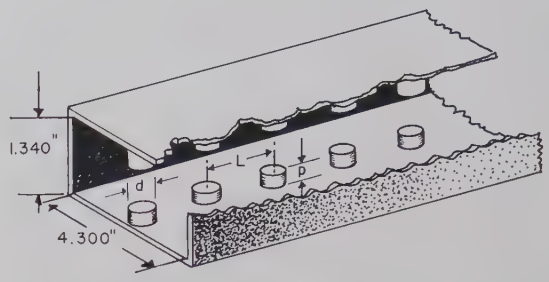


Fig. 1—Periodic structure.

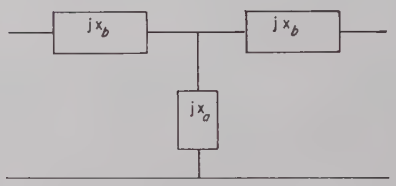


Fig. 2—Equivalent circuit of post.

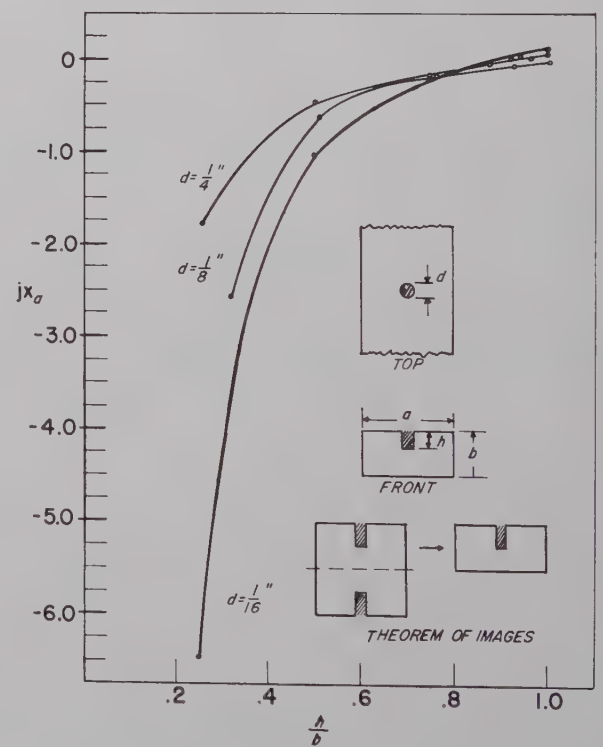


Fig. 3—Shunt reactance of post. (*Experimental curves from Marcuvitz.*[2])

* Received by the PGMTT, April 10, 1961; revised manuscript received, May 31, 1961.

† Electromagnetic Rad. Lab., AF Research Labs., Office of Aerospace Research, USAF, Bedford, Mass.

[1] M. Suzuki, "Circuit Parameters of a Tuning Post in a Rectangular Waveguide and its Applications," AF Cambridge Research Center, Bedford, Mass., AFCRC-TN-57-764 [Rept. R-591-57, PIB, 519, AFCRC Contract No. AF 19(604)-2031]; July, 1957.

[2] N. Marcuvitz, "Waveguide Handbook," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 10, p. 271; 1947.

it behaves as a short circuit) and then becomes inductive. It can be shown from the theorem of images (see Fig. 3) that the shunt reactance of the double post being investigated is equivalent to that of a single post in a waveguide of half the height.

The effective guide wavelength λ_g' and characteristic impedance Z' of a periodic structure are related to the guide wavelength λ_g and characteristic impedance Z of the unloaded waveguide as follows:

$$\cos\frac{2\pi}{\lambda_g'}L = \cos\frac{2\pi}{\lambda_g}L - \frac{1}{2x_a}\sin\frac{2\pi}{\lambda_g}L, \tag{1}$$

$$\frac{Z'}{Z} = \frac{\tan\dfrac{\pi}{\lambda_g}L}{\tan\dfrac{\pi}{\lambda_g'}L}, \tag{2}$$

where L is the distance between posts, and x_a is the normalized shunt reactance of the posts.[3] (It is assumed that $x_b=0$.)

Eqs. (1) and (2) are valid only when the posts are far enough apart so that the coupling between them is negligible. At a fixed frequency (λ_g=constant), λ_g' is a function of x_a and L. For a fixed post separation of L/λ_g, as x_a is made less negative, cos $(2\pi L/\lambda_g')$ approaches -1. When cos $(2\pi L/\lambda_g')$ is greater than -1, it becomes indeterminant and this corresponds to a cutoff region in the waveguide. For a fixed L/λ_g, therefore, λ_g' can be decreased by only a limited amount regardless how small x_a is made. Upon examining the limiting condition where

$$\cos\frac{2\pi L}{\lambda_g'} = -1,$$

however, it can be seen that $\lambda_g'=2L$ and so λ_g' can be made smaller by decreasing the distance between posts. When this distance becomes too small, the coupling between posts cannot be neglected, (1) is no longer valid, and λ_g' cannot be further decreased.

When x_a is made inductive (positive), cos $(2\pi L/\lambda_g')$ approaches $+1$ and λ_g' approaches infinity. If the posts become too inductive

$$[\cos(2\pi L/\lambda_g') > 1],$$

the TE_{10} mode does not propagate.

The behavior of the periodic structure of cylindrical posts can be summarized as follows: When the post penetration is small the post can be represented by a shunt capacitance, and the guide wavelength λ_g' of the periodic structure is less than the guide wavelength λ_g of the unloaded waveguide. As the post penetration is further increased, λ_g' is further decreased until a cutoff region is reached and propagation of the TE_{10} mode ceases. With a further increase in post penetration, the shunt reactance passes through a resonance ($x_a=0$) and then becomes inductive. When the inductive reactance increases to a large enough value such that cos $(2\pi L/\lambda_g')<+1$, propagation resumes and λ_g' decreases from infinity down to a value larger than λ_g as the post is extended all the way across the waveguide.

[3] J. Brown, "The design of metallic delay dielectrics," *Proc. IEE*, vol. 97, pt. 3, p. 45; January, 1950.

Experimental Procedure

The guide wavelength λ_g' of a periodic structure of known length can be determined by measuring the phase of a traveling wave at the beginning and end of the structure if the guide wavelength λ_g of the unloaded waveguide is known. The phase measurement is made using the experimental setup shown in Fig. 4. The input signal is divided into two parts; one part travels through the periodic structure under test, and the other part travels through a slotted line that is used as a reference phase line. To obtain precise data it is necessary that both signals be traveling waves. Each waveguide is therefore terminated in a matched load. The input signals to arms 1 and 2 of the magic T are adjusted so that they are approximately equal in amplitude. A sharp null will appear at output arm 3 when the input signals are 180° out of phase and arm 4 is terminated in a matched load. The phase difference between two signals is determined from the distance that the probe on the phase line (slotted line No. 1) has to be moved to produce a null at output arm 3.

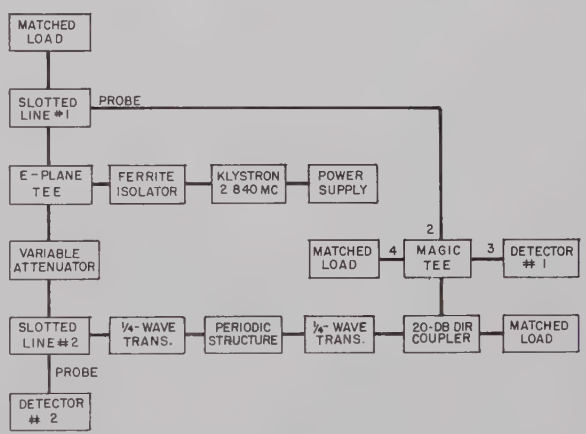


Fig. 4—Block diagram of equipment.

The periodic structure consists of a uniform section of waveguide with adjustable posts and behaves as a uniform waveguide with a propagation constant and characteristic impedance different from that of the unloaded waveguide. The phase of the signal from the directional coupler at the end of the periodic structure is first measured with the posts at zero depth. The post penetration is then slightly increased and the measurement is repeated. The periodic structure has an

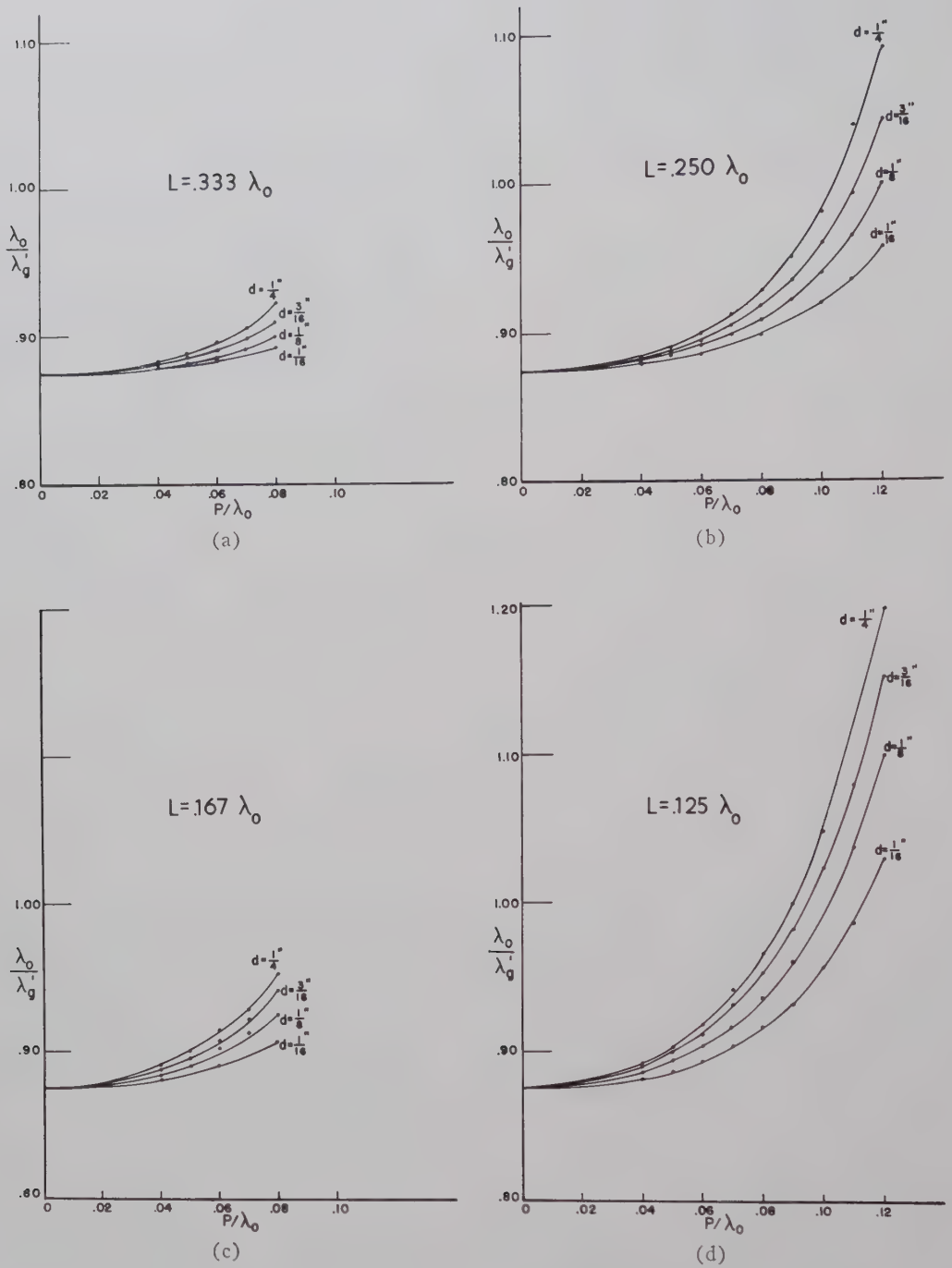


Fig. 5—Guide wavelength vs post depth for constant post spacing.

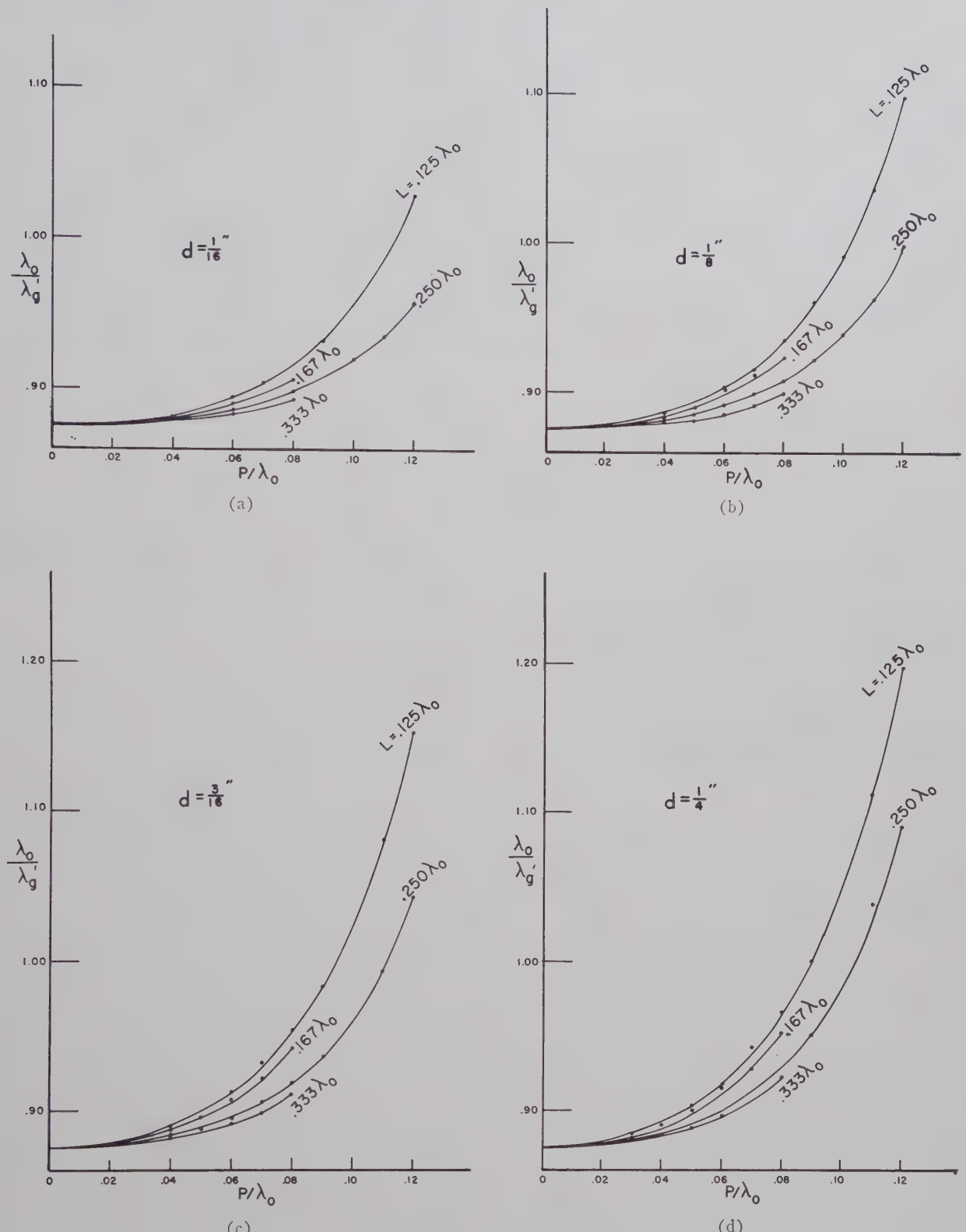


Fig. 6—Guide wavelength vs post depth for constant post diameter.

electrical length of 630° ($1.75\lambda_g$) with zero post penetration. As the post depth is increased, the shift in phase increases. Since this corresponds to an increase in the electrical length of the structure, the guide wavelength is decreased. The ratio of freespace wavelength λ_0 to the guide wavelength λ_g of the unloaded structure is 0.875. The ratio of λ_0 to λ_g' is

$$\frac{\lambda_0}{\lambda_g'} = 0.875\left[1 + \frac{\Delta\theta}{630°}\right],$$

where $\Delta\theta$ is the change in phase in degrees due to the presence of the posts.

Measurement of the input VSWR of the periodic structure checks the relationship.

$$\frac{Z'}{Z} = \frac{\tan\frac{\pi}{\lambda_g}L}{\tan\frac{\pi}{\lambda_g'}L}.$$

This is possible because the input VSWR is composed of reflections from the interface from Z to Z' and the interface from Z' to Z. Since the characteristic impedances are assumed real, the VSWR that is due to each interface is R/R' for $R>R'$. The resultant input VSWR can vary from a minimum of 1.0 to a maximum of $(R/R')^2$, depending on the electrical distance between interfaces. It is 1.0 when the interfaces are an even number of quarter-wavelengths apart and $(R/R')^2$ when they are an odd number of quarter-wavelengths apart. Due to the reflections at the interfaces, there is a slight difference between the phase angle of the transmission coefficient that is measured and the electrical length of the periodic section. Since the reflection coefficients are small, the error introduced is negligible.

Discussion of Results

The guide wavelength λ_g' of the periodic structure shown in Fig. 1 was measured as a function of post diameter d, penetration p, and spacing L, at a frequency of 2840 Mc ($\lambda_0=4.159$ inches). In Figs. 5 and 6, λ_0/λ_g' is plotted as a function of p/λ_0. Each set of curves in Fig. 5 corresponds to a fixed post spacing L. Each set of curves in Fig. 6 corresponds to a fixed post diameter d. It can be seen that λ_0/λ_g' increases slowly with an increase in post diameter and depth and a decrease in post spacing. These results agree very well with those expected from the qualitative discussion that was presented earlier. Since this investigation was conducted primarily in the region where $\lambda_g'\cong\lambda_0$, the actual cutoff conditions of the periodic structure were not established; but the fact that the slopes of the curves become very steep for large post diameter and penetration indicates that a cutoff condition is rapidly being approached.

The results of the input VSWR measurements were found to be consistent with those expected from (2). The assumption that this periodic structure can be treated as a uniform waveguide with a real characteristic impedance is therefore valid.

Acknowledgment

The author wishes to thank Leon Dorr and Robert Beyer for making the experimental measurements. Special acknowledgment is due Walter Rotman for his helpful suggestions and criticism.

Modes in Rectangular Guides Loaded with a Transversely Magnetized Slab of Ferrite away from the Side Walls*

G. BARZILAI†, SENIOR MEMBER, IRE, AND G. GEROSA†, ASSOCIATE, IRE

Summary—**The characteristic equation describing the general modal spectrum for a rectangular guide partially filled with a slab of ferrite transversely magnetized and situated away from the side walls is derived. This equation is numerically solved for particular cases and for modes of zero and first order with respect to the dependence along the direction of the dc magnetic field. Some experiments to verify the theoretical results are presented and show good agreement with the theory.**

I. Introduction

THE authors in a previous paper[1] have discussed the modal spectrum in rectangular guides completely filled with transversely magnetized ferrite; in a successive paper[2] they have discussed the modal spectrum in rectangular guides loaded with a slab of transversely magnetized ferrite against one side wall.

The purpose of this work is to report on further theoretical study on a more general structure of a rectangular guide loaded with a slab of transversely magnetized ferrite away from the side walls and to describe some experiments carried out to verify the theory. The only analyses available up to date for this structure are, to our knowledge, relative to modes with no dependence along the direction of the dc magnetic field.[3–6]

The method of solution we have used is similar to the one already discussed,[1] but the derivation of the characteristic equation is very involved and the relative numerical computations, even by using a medium speed electronic computer, require considerable length of time. We have therefore limited our numerical investigation to some cases which we thought to be of interest.

II. The Characteristic Equation

Let us refer to the parallel plate guide represented in Fig. 1. We shall assume the walls to be perfectly conducting.

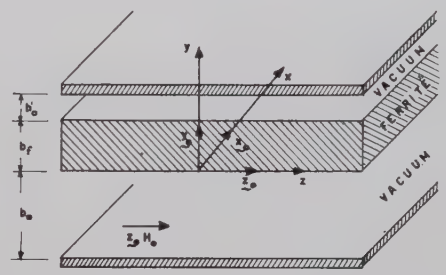


Fig. 1—Geometry of the parallel plate guide partially filled with magnetized ferrite.

Let us assume the dc magnetic field, of sufficient intensity H_0 to saturate the ferrite, to be directed along the z axis. We shall assume for the ferrite region a scalar dielectric constant $\epsilon_0\,\epsilon$ and a magnetic tensor permeability $\underline{\mu}$ given by the following expression (time dependence exp $[j\omega t]$):

$$\underline{\mu} = \mu_0 \begin{vmatrix} \mu_1 & j\mu_2 & 0 \\ -j\mu_2 & \mu_1 & 0 \\ 0 & 0 & 1 \end{vmatrix},$$

where

$$\mu_1 = 1 + \frac{\rho}{1-\tau^2}\,; \quad \mu_2 = \frac{\tau\rho}{1-\tau^2}\,; \quad \rho = \frac{M_0}{\mu_0 H_0}\,; \quad \tau = \frac{\omega}{\omega_0},$$

and M_0 is the intensity of the saturation magnetization, ω and $\omega_0 = -\gamma H_0$ are the applied and the resonant circular frequencies, γ is the gyromagnetic ratio for the electron, μ_0 and ϵ_0 are the permeability and the dielectric constant of the vacuum.

As discussed in detail in our previous work,[1] to construct our modal solution we shall consider for the three

* Manuscript received by the PGMTT, February 20, 1961; revised manuscript received, June 15, 1961. This paper was presented at the 13th General Assembly of URSI, London, Eng., September 5–15, 1960. The theoretical investigation reported herein was sponsored by the U. S. Air Force under Contract No. AF-61 (052)-101. The experimental part was made possible by the support of the Consiglio Nazionale delle Ricerche of Italy.

† Istituto di Elettronica dell Università di Roma, Rome Italy.

[1] G. Barzilai and G. Gerosa, "Modes in rectangular guides filled with magnetized ferrite," *L'Onde électrique, 38e Année*, N° 376ter, Suppl. Spécial—Congrès International Circuits et Antennes Hyperfréquences, Paris, France; pp. 612–617; October 21–26, 1957.

[2] G. Barzilai and G. Gerosa, "Modes in rectangular guides partially filled with transversely magnetized ferrite," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-7, pp. S 471–S 474; December, 1959. Also see Istituto Electrotecnico dell'Università di Roma, Tech. Note No. 1, Contract No. AF 61(052)-101; June 3, 1959.

[3] M. L. Kales, H. N. Chait and N. G. Sakiotis, "A nonreciprocal microwave component," *J. Appl. Phys.*, vol. 24, pp. 816–817; June, 1953. Erratum: p. 1528; December, 1953.

[4] B. Lax, K. J. Button and L. M. Roth, "Ferrite phase shifters in rectangular waveguide," *J. Appl. Phys.*, vol. 25, pp. 1413–1421; November, 1954.

[5] B. Lax and K. J. Button, "Theory of ferrites in rectangular waveguide," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. 4, pp. 531–537; July, 1956.

[6] A. L. Mikaelyan, "Magneto-optic phenomena in a rectangular waveguide containing a ferrite plate," *Izvest. Akad. Nauk. USSR, Otdel. Tekh. Nauk.*, No. 3, pp. 139–144; 1955.

regions of width b_f, b_0 and b_0' fields having spatial dependence of the form:

$$\exp\left[j(k_x x + k_y y + k_z z)\right]. \tag{1}$$

We shall add to the propagation constants k_x, k_y, k_z a subscript f or 0 in order to refer to the ferrite or to the vacuum regions.

The propagation constants will be measured in units of $\omega\sqrt{\mu_0\epsilon_0}$ and consequently lengths will be measured assuming as unit $1\ \omega\sqrt{\mu_0\epsilon_0}$.

In order to satisfy Maxwell's equations, we must have in the ferrite region

$$\mu_1 t^4 + \left[(\mu_1 + 1)k_{zf}^2 - \epsilon(\mu_1^2 - \mu_2^2 + \mu_1)\right]t^2 + k_{zf}^4 - 2\mu_1\epsilon k_{zf}^2 + \epsilon^2(\mu_1^2 - \mu_2^2) = 0, \tag{2}$$

where

$$t^2 = k_{xf}^2 + k_{yf}^2, \tag{3}$$

and in the vacuum regions

$$k_{x0}^2 + k_{y0}^2 + k_{z0}^2 = 1. \tag{4}$$

Each modal solution of our problem is labelled by a pair of values k_z, k_x. In order to satisfy the boundary conditions at the interfaces we must have: $k_{zf} = k_{z0} = k_z$ and $k_{xf} = k_{x0} = k_x$.

For the ferrite region (2) yields, for a given k_z, two values of t^2: t_1^2 and t_2^2. From (3) we obtain therefore two values of k_{yf}^2: k_{y1}^2 and k_{y2}^2, for a given k_x. Consequently we express the field in the ferrite region as a superposition of four fields of the form (1) having the same k_x and k_z and arbitrary amplitudes.

In the vacuum regions for a given pair k_z, k_x we have only one value of k_{y0}^2, but, since a general field can be expressed as a superposition of TE and TM waves, for each vacuum region we shall have again four arbitrary amplitudes.

By imposing the tangential component of the electric field to be zero at the metallic boundaries and the tangential component of both the electric and magnetic fields to be continuous at the interfaces between vacuum and ferrite, we obtain a system of twelve linear algebraic homogeneous equations in the twelve unknown amplitudes. By setting equal to zero the determinant of the coefficients, we obtain the following characteristic equation:

$$\sum_{(i,k)} \sum_{(l,m)} (-1)^{\cdots} \Gamma_{(i,k)(l,m)}(b_f) \cdot X_{i,k}(b_0) X_{l,m}(-b_0') = 0, \tag{5}$$

where the index pairs i, k and l, m can assume the values 1,2; 1,3; 1,4; 2,3; 2,4; 3,4 and $(i, k)_c$ represents the pair which together with i, k complete the set of first four integral numbers; and similarly for $(l, m)_c$.

The functions X and Y have the following expressions:

$$X_{i,k}(b_0) = c_{i1}c_{k2} - c_{i2}c_{k1},$$

where

$$c_{11} = k_z(1 - k_x^2)k_{y0}\sin k_{y0}b_0 \qquad c_{12} = 0$$
$$c_{21} = -k_x k_z^2 k_{y0}\sin k_{y0}b_0 \qquad c_{22} = -k_{y0}\sin k_{y0}b_0$$
$$c_{31} = 0 \qquad c_{32} = -(1 - k_x^2)\cos k_{y0}b_0$$
$$c_{41} = -k_z k_{y0}^2\cos k_{y0}b_0 \qquad c_{42} = k_x k_z \cos k_{y0}b_0$$

and

$$\begin{aligned}\Gamma_{(i,k)(l,m)}(b_f) = {} & -k_{y1}k_{y2}(C_{ik,11}C_{lm,22} + C_{ik,22}C_{lm,11}) \\ & + \sin k_{y1}b_f \sin k_{y2}b_f(A_{ik,12}A_{lm,12} \\ & + k_{y1}^2 k_{y2}^2 B_{ik,12}B_{lm,12} + k_{y1}^2 C_{ik,21}C_{lm,21} \\ & + k_{y2}^2 C_{ik,12}C_{lm,12}) \\ & + k_{y1}k_{y2}\cos k_{y1}b_f \cos k_{y2}b_f(B_{ik,12}A_{lm,12} \\ & + A_{ik,12}B_{lm,12} + C_{ik,21}C_{lm,12} + C_{ik,12}C_{lm,21}) \\ & + k_{y1}\cos k_{y1}b_f \sin k_{y2}b_f[C_{ik,21}A_{lm,12} \\ & - A_{ik,12}C_{lm,21} + k_{y2}^2(C_{ik,12}B_{lm,12} \\ & - B_{ik,12}C_{lm,12})] \\ & + k_{y2}\sin k_{y1}b_f \cos k_{y2}b_f \cdot [A_{ik,12}C_{lm,12} \\ & - C_{ik,12}A_{lm,12} + k_{y1}^2(B_{ik,12}C_{lm,21} \\ & - C_{ik,21}B_{lm,12})],\end{aligned}$$

where

$$\begin{aligned}A_{ik,12} &= a_{i1}a_{k2} - a_{k1}a_{i2} \\ B_{ik,12} &= b_{i1}b_{k2} - b_{k1}b_{i2} \\ C_{ik,11} &= a_{i1}b_{k1} - a_{k1}b_{i1} \\ C_{ik,22} &= a_{i2}b_{k2} - a_{k2}b_{i2} \\ C_{ik,21} &= a_{i2}b_{k1} - a_{k2}b_{i1} \\ C_{ik,12} &= a_{i1}b_{k2} - a_{k1}b_{i2}\end{aligned}$$

and similarly for l, m, and:

$$a_{11} = \mu_2 k_z(\epsilon - k_x^2) = a_{12}$$
$$b_{11} = -(\mu_1 - 1)k_z k_x = b_{12}$$
$$a_{21} = -\mu_2(\epsilon - t_1^2)k_x \qquad a_{22} = -\mu_2(\epsilon - t_2^2)k_x$$
$$b_{21} = k_z^2 - \mu_1(\epsilon - t_1^2) \qquad b_{22} = k_z^2 - \mu_1(\epsilon - t_2^2)$$
$$a_{31} = k_x^2[k_x^2 - (\epsilon - t_1^2)] - \epsilon[k_z^2 - \mu_1(\epsilon - t_1^2)] \qquad a_{32} = k_x^2[k_z^2 - (\epsilon - t_2^2)] - \epsilon[k_z^2 - \mu_1(\epsilon - t_2^2)]$$
$$b_{31} = 0 = b_{32}$$
$$a_{41} = -k_z(\mu_1\epsilon - k_z^2 - t_1^2)k_x \qquad a_{42} = -k_z(\mu_1\epsilon - k_z^2 - t_2^2)k_x$$
$$b_{41} = -k_z\mu_2\epsilon = b_{42}.$$

It can be seen that (5) contains only even powers of k_z.

It can be verified that the solution possesses reflection symmetry along the z direction. Therefore a solution for a rectangular guide of height a can be obtained by choosing

$$k_z = \frac{m\pi}{a}, \qquad m = 0, 1, 2 \cdots \tag{6}$$

For $m=0$, (5) breaks down into two equations corresponding to TE and TM modes. TM modes cannot exist in the rectangular guide obtained by closing the structure of Fig. 1 with two metallic walls normal to the z-axis. TE modes can exist and the relative characteristic equation is the same as the one given by others. From $m\neq 0$, (5) does not in general break down into two equations and the resulting electromagnetic field cannot be resolved into TE and TM modes.

The solution does not possess reflection symmetry along the x axis. In addition, (5) contains in general odd powers of k_x, and the solution therefore is not reciprocal, *i.e.*, a solution $+k_x$ does not necessarily imply the solution $-k_x$.

It should be noted that obvious solutions of (5) are $k_{y1}=0$; $k_{y2}=0$; $k_{y0}=0$. These solutions however correspond to zero amplitude fields. It can be verified that

$$k_x^2 = \epsilon \frac{k_z^2 - \mu_1(\epsilon - t_{1,2}^2)}{k_z^2 - (\epsilon - t_{1,2}^2)}$$

are always solutions of (5), but these solutions also correspond to zero amplitude fields.

We note that by interchanging b_0 with b_0' and k_x with $-k_x$, (5) remains unchanged as it follows from symmetry considerations.

To discuss (5), it is convenient to find the asymptotic behavior of some of the solutions when $b_0'\to 0$. More exactly we shall assume b_f and b_0 finite and by letting $b_0'\to 0$ we shall look for solutions with real k_x such that $|k_x|\to\infty$. From (5) we obtain

$$\tanh |k_x| b_0' = \frac{\mu_2^2 - \mu_1(\mu_1+1) - \dfrac{k_x}{|k_x|}\mu_2}{\mu_1 + 1 - \dfrac{k_x}{|k_x|}\mu_2} . \qquad (7)$$

From (7) it is apparent that asymptotic solutions are possible only when the second member of the equation is positive. In such a case the asymptotic solutions are represented by hyperbolas. It should also be noted that (7) is independent of k_z and therefore the asymptotic solutions considered are the same for any k_z, *i.e.*, in the case of the rectangular guide for any m.

III. Numerical Analysis

There are several parameters which determine the solutions of (5), namely: quantities characteristic of the ferrite medium, *i.e.*, M_0 and ϵ; quantities describing the structure, *i.e.*, a, b, b_f and b_0' ($b_0=b-b_f-b_0'$); impressed quantities, *i.e.*, ω and H_0. The three quantities M_0, ω and H_0 enter our problem through the adimensional parameters ρ and τ, so that the actual parameters to be considered are ρ, ϵ, τ, a, b, b_f and b_0'.

When a set of these parameters has been chosen, (5) in virtue of (2), (3) and (4) becomes a relation between k_x and k_z^2. We shall fix k_z by means of (6), which allows us to label our modal solutions according to the integral values given to m. We shall call, therefore, modes of zero order those corresponding to $k_z=0$, and modes of first order those corresponding to $k_z=\pi/a$, etc.

For a given mode order, (5) determines the relative propagation constants k_x. In what follows we shall only look for real values of k_x, *i.e.*, for unattenuated propagating modes.

With respect to the field configuration in the cross section of the guide the mode order determines the z dependence. The y dependence is determined by the values of k_{y1}, k_{y2} and k_{y0}, which can assume real or purely imaginary values. In fact, by assuming k_z and k_x real, (4) states that k_{y0}^2 is real and (2), once solved with respect to t^2, shows that t_1^2 and t_2^2 are both real.

The solutions of (5) can be classified by dividing the field of variability of τ into six regions, exactly in the same manner as we have done in our previous work[2] relative to the slab of ferrite against one side wall.

We have numerically solved (5) for three cases which we thought to be of interest. Two of such cases are those corresponding to Figs. 5 and 6 of the quoted paper.[2]

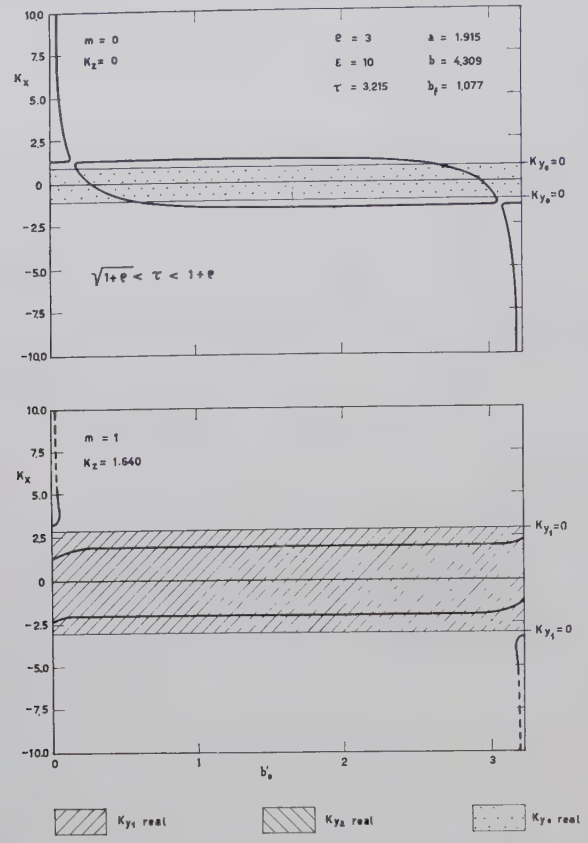
The results of the numerical analysis are recorded in the diagram of Figs. 2–4, which show k_x vs b_0' for the values of the parameters indicated in the figures and for modes of zero and first order.

In the diagrams of Figs. 2–4 the various zones delimited by the straight lines $k_{y1}=0$, $k_{y2}=0$ and $k_{y0}=0$ have been shaded in different ways in order to recognize at a glance when k_{y1}, k_{y2}, and k_{y0} are real or imaginary. It is understood that when one of the three typical shadings indicated is present the relative k_y is real and when it is not it is imaginary. For instance, when no shading exists the three k_y's are all imaginary; when all three shadings are present the three k_y's are all real, and so on. It should be noted however that for modes of zero order k_{y1} is associated with TM modes, which have zero amplitude in the rectangular guide. The shading relative to k_{y1} has been therefore omitted in the diagrams relative to zero order modes.

The most interesting feature appearing from the diagrams is the fact that for suitable dimensions of the guide and of the ferrite slab unidirectional propagation exists for $b_0'=0$ and, as soon as the slab of ferrite is moved sufficiently away from the wall, for zero and first order modes no propagation can occur in either direction. It is reasonable to assume that higher order modes exhibit the same behavior. Experimental evidences described in the next section seem to agree with this theoretical result.

In Fig. 4 we have indicated with b_{0c}' the cutoff distance for modes of zero order. The cutoff distance for modes of first order is practically the same.

With reference to the zero order modes of Fig. 4 we have investigated the sign of the group velocity for the region where there are two propagating modes. We note



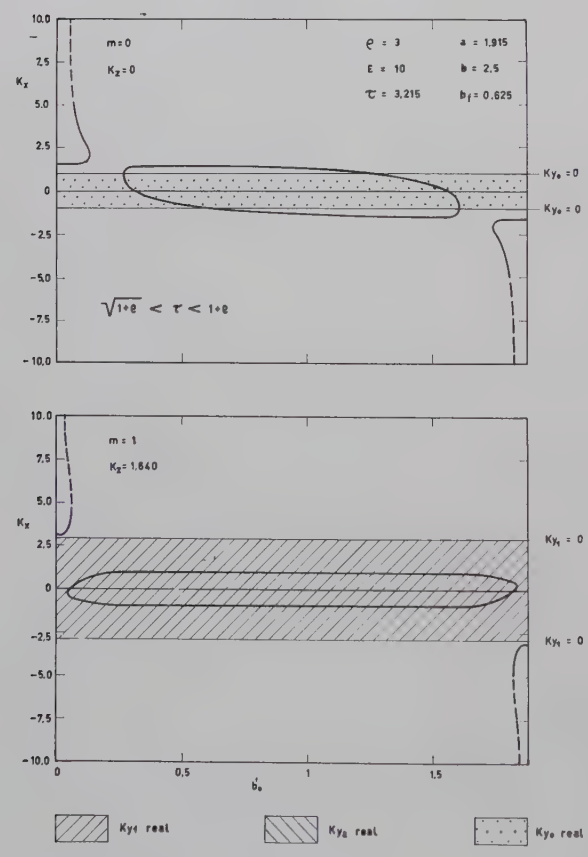


Fig. 3—The same as Fig. 2, except for

$$\frac{b}{\omega\sqrt{\mu_0\epsilon_0}} = 13.25 \text{ mm}; \qquad \frac{b_f}{\omega\sqrt{\mu_0\epsilon_0}} = 3.30 \text{ mm}.$$

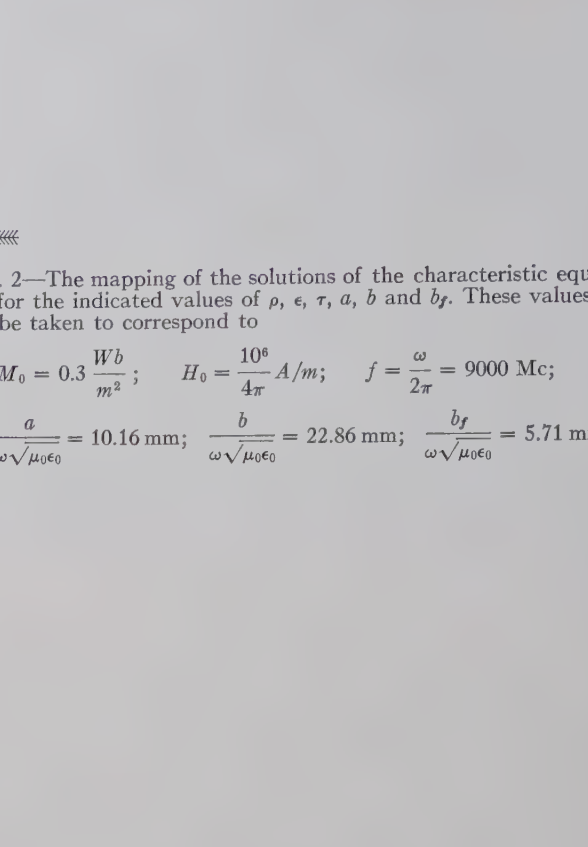


Fig. 2—The mapping of the solutions of the characteristic equation for the indicated values of ρ, ϵ, τ, a, b and b_f. These values may be taken to correspond to

$$M_0 = 0.3\,\frac{Wb}{m^2}; \qquad H_0 = \frac{10^6}{4\pi} A/m; \qquad f = \frac{\omega}{2\pi} = 9000 \text{ Mc};$$

$$\frac{a}{\omega\sqrt{\mu_0\epsilon_0}} = 10.16 \text{ mm}; \quad \frac{b}{\omega\sqrt{\mu_0\epsilon_0}} = 22.86 \text{ mm}; \quad \frac{b_f}{\omega\sqrt{\mu_0\epsilon_0}} = 5.71 \text{ mm}.$$

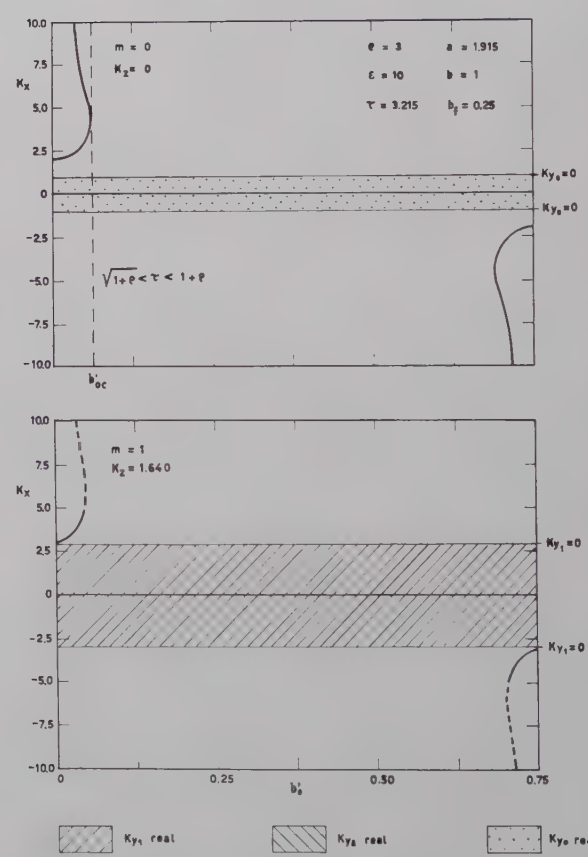
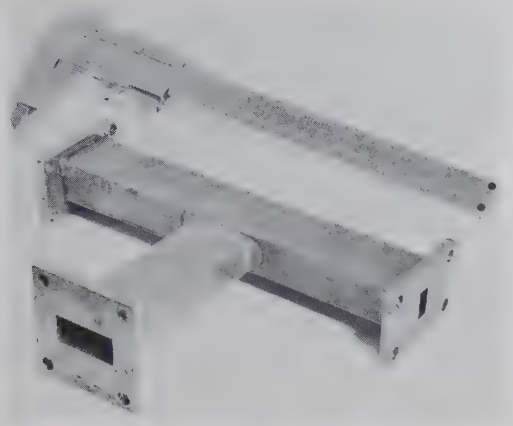
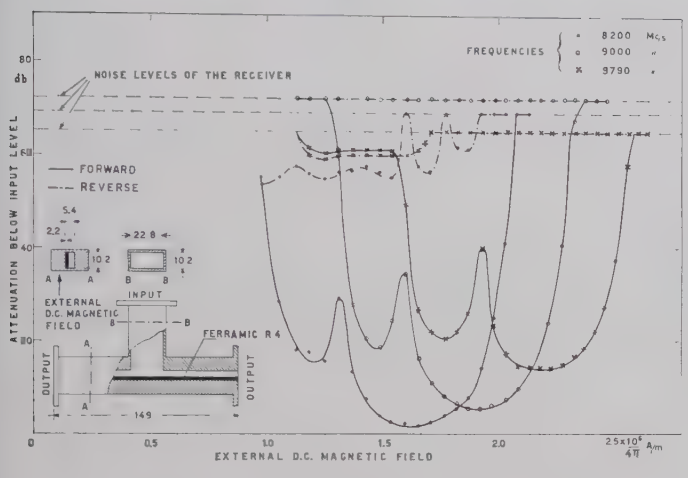


Fig. 4—The same as Fig. 2, except for

$$\frac{b}{\omega\sqrt{\mu_0\epsilon_0}} = 5.30 \text{ mm}; \qquad \frac{b_f}{\omega\sqrt{\mu_0\epsilon_0}} = 1.32 \text{ mm}.$$

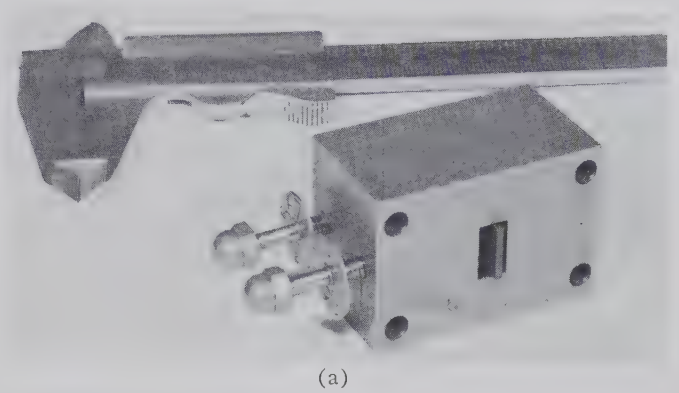


(a)

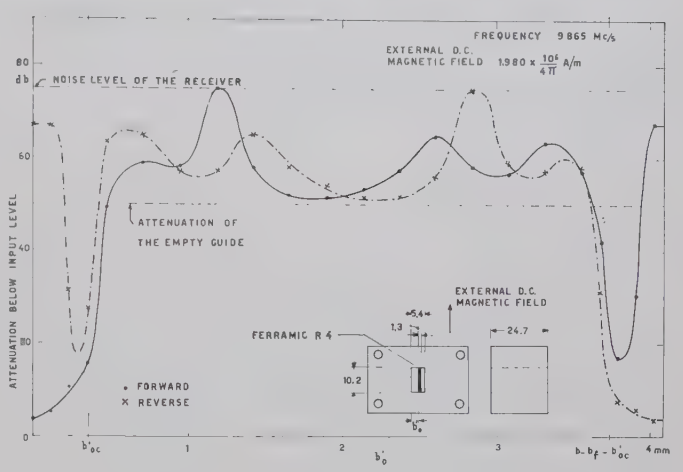


(b)

Fig. 5—(a) T-shaped waveguide built to verify the unidirectional character of a guide of suitable dimensions loaded with magnetized ferrite. (b) Experimental curves of the attenuation below the input level vs the external dc magnetic field for the structure and the frequencies indicated. All the dimensions are in millimeters.



(a)



(b)

Fig. 6—(a) Waveguide section built to investigate the behavior of a guide loaded with a slab of magnetized ferrite. (b) Experimental curves of the attenuation below the input level vs b_0' for the values of the frequency and the external dc magnetic field indicated. All the dimensions are in millimeters.

that, because of the normalization assumed, the group velocity u_g is given by

$$u_g = \frac{1}{\sqrt{\mu_0 \epsilon_0}\, \dfrac{d(\tau k_x)}{d\tau}} .$$

By using the asymptotic expression (7) it is easily seen that by increasing the frequency, *i.e.*, τ, τk_x decreases and vice versa. On the other hand for $b_0' = 0$ we have computed τk_x for two values of the frequency slightly above and below the frequency relative to Fig. 4 and we have found that the group velocity is positive. By using these results we can conclude that the two modes corresponding to the same value of b_0' have group velocity of opposite sign. It seems reasonable to extend the validity of this result to modes of the first order.

IV. Experimental

The experiments carried out had two different aims: 1) to prove the unidirectionality of a structure corresponding to the case of Fig. 4 for $b_0' = 0$; 2) to verify the general behavior of a structure corresponding to the case of Fig. 4 when b_0' is varied.

For the experiment 1) we have built a T-shaped waveguide [shown in Fig. 5(a)] whose dimensions are given in Fig. 5(b). By sending the RF energy in the input arm we have recorded in Fig. 5(b) the attenuation for both the output arms below the input level vs the external dc magnetic field.[7] This has been done for the three different frequencies indicated in the figure. From the experimental results of Fig. 5(b) the unidirectional character of the structure considered is apparent for a wide band of frequencies.

For the experiment 2) we have built a guide section [shown in Fig. 6(a)] whose dimensions are given in Fig. 6(b). By means of insulating rods and screws the slab of ferrite could be moved parallel to itself inside the guide. By sending energy in the input arm we have recorded in Fig. 6(b) the attenuation below the input level at the output arm vs the distance of the slab of ferrite from one of the walls, for the two opposite values of the dc magnetic field, which correspond to

[7] Note that the dc magnetic field assumed for the theoretical calculations is the internal one, which in the experiments is not uniform and its average is smaller than the external.

forward and reverse propagation. In the figure it is shown the attenuation for the empty guide, *i.e.*, for the same guide without the slab of ferrite.

Let's follow the forward propagation experimental curve from left to right. When the slab of ferrite is against the left side wall the attenuation is at a minimum. By moving the slab of ferrite away from the wall, after a region of low attenuation, the signal goes very rapidly below the level of the attenuation of the empty guide. This is justified by the theoretical results of Fig. 4 and by the considerations about the group velocity. In fact, beyond the distance b_{0c}' no propagation can exist and below such a distance always exists a propagating mode with group velocity in the forward direction.

Moving the slab further away from the left side wall, the signal remains at a level below the empty guide attenuation level until we reach a distance approximately equal to $b-b_f-b_{0c}'$ when the attenuation begins to decrease and then, after it has reached a minimum, increases again above the empty guide attenuation. This last behavior is easily explained by the theoretical results, since beyond the distance $b-b_f-b_{0c}'$, energy begins to pass in the forward direction through modes having positive group velocity. However, since these modes have propagation constants going to $-\infty$ as the slab of ferrite approaches the right-hand side wall and cannot therefore allow propagation in the limit, the minimum of the attenuation experimentally found is explained.

From the preceding discussion we can conclude that there is good agreement between theory and experiments.

Higher-Order Evaluation of Electromagnetic Diffraction by Circular Disks*

W. H. EGGIMANN†

Summary—**The problem of the diffraction of an arbitrary electromagnetic field by a circular perfectly-conducting disk has been solved by using a series representation in powers of $k=2\pi/\lambda$ and the rectangular disk coordinates. The surface current density is given in terms of the field and its derivatives at the center of the disk. General expressions for the electric- and magnetic-dipole moments, the far-field and the scattering coefficient for the case of a plane wave at arbitrary incidence are presented. The calculations agree with results published by other authors. A bibliography of the most recent publications on this problem is included.**

I. Introduction

THE problem of the diffraction by a circular conducting disk (or the complementary problem for a circular aperture in an infinite plane conducting screen) has occupied many workers in the field of diffraction theory. The problem can be formulated as follows:

1) the electromagnetic field has to obey Maxwell's equations,
2) the boundary conditions on the surface of the disk have to be fufilled, *e.g.*, for a perfectly-conducting disk the tangential electric field must vanish,
3) the edge conditions [48] at the rim of the disk have to be obeyed; they require that the field energy remains finite, or that the energy density has to be integrable over any finite space. This leads to the requirement that the normal component of the electric field increases not faster than $(1/r)^{1/2}$ where r is the distance from the edge,
4) Sommerfeld's radiation conditions [47] have to be fulfilled.

In this paper a power-series solution in (ka) valid for the small disk problem ($a<\lambda/2\pi$, where $a=$disk radius, $\lambda=$free-space wavelength) and an arbitrary incident field is given. It is essentially an extension of a procedure described by Bouwkamp [45]. The surface current density on the disk up to the third-order approximation in (ka) is calculated in terms of the electromagnetic field and its derivatives at the center of the disk. From these results expressions for the induced electric and magnetic dipole moments and the far-zone fields are derived. The scattering coefficient for a plane wave at arbitrary incidence has been calculated in agreement with formulas given by Lur'e [19] and Kuritsyn [20]. The essential advantage of the expressions obtained in this paper is that they can be used for any primary field. This is important in the case where interaction between several disks is considered. If the spacing between the disks is not large compared with the wavelength, the interaction fields cannot be approximated by a plane wave and the interaction be-

* Received by the PGMTT, April 10, 1961; revised manuscript received, June 8, 1961. This work has been sponsored by the Electronics Res. Dir. of the AFCRC under contract no. AF 19(604)-3887.

† Dept. of Elec. Engrg., Case Inst. of Tech., Cleveland, Ohio.

between the induced dipole moments and higher-order multipole moments have to be taken into account. Some of this work will be reported at a later date.

II. Power-Series Solution

General Formulation

We shall consider a perfectly-conducting circular disk of radius a with its axis along the z direction as in Fig. 1. The magnetic vector potential $\boldsymbol{A}(\boldsymbol{R})$ for the diffracted field is given by

$$\boldsymbol{A}(\boldsymbol{R}) = \frac{\mu_0}{4\pi}\int_D \boldsymbol{J}(\boldsymbol{p}')\frac{e^{-jkr}}{r}\,dS' \tag{1}$$

where $\boldsymbol{J}(\boldsymbol{p}')$ is the electric surface current density and the integration is carried out over the disk D. $\boldsymbol{R}$ and $\boldsymbol{p}'$ are the coordinates of the field point $P(x, y, z)$ and the surface element dS' respectively, while $\boldsymbol{r}$ is the radius vector from dS' to $P(x, y, z)$. The time dependence $e^{j\omega t}$ is omitted throughout. The scattered field is found from the following relations:

$$\boldsymbol{H} = \frac{1}{\mu_0}\nabla \times \boldsymbol{A} \tag{2}$$

$$\boldsymbol{E} = -j\omega\boldsymbol{A} + \frac{\nabla\nabla\cdot\boldsymbol{A}}{j\omega\mu_0\epsilon_0}. \tag{3}$$

The boundary conditions are satisfied if

$$\begin{aligned} E_x(\boldsymbol{R}) &= -E_x{}^i(\boldsymbol{R}) \\ E_y(\boldsymbol{R}) &= -E_y{}^i(\boldsymbol{R}) \quad \text{on the disk} \\ H_z(\boldsymbol{R}) &= -H_z{}^i(\boldsymbol{R}) \end{aligned} \tag{4}$$

where the superscript i indicates the incident field. We now express (2) and (3) in rectangular coordinates and combine them with (4)

$$H_z = \frac{1}{\mu_0}\left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) = -H_z{}^i \tag{5}$$

$$\begin{aligned} E_x &= -j\omega A_x + \frac{1}{j\omega\mu_0\epsilon_0}\frac{\partial}{\partial x}\left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y}\right) \\ &= -j\omega A_x + \frac{1}{j\omega\mu_0\epsilon_0}\left[\frac{\partial^2 A_x}{\partial x^2} + \frac{\partial}{\partial y}\left(\frac{\partial A_x}{\partial y} - \mu_0 H_z{}^i\right)\right] \\ &= -j\omega A_x + \frac{1}{j\omega\epsilon_0\mu_0}\left[\nabla_{xy}{}^2 A_x - \mu_0\frac{\partial H_y{}^i}{\partial z} - j\omega\mu_0\epsilon_0 E_x{}^i\right] \\ &= -E_x{}^i \end{aligned} \tag{6}$$

$\nabla_{xy}{}^2 = \dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2}$ is defined as the two-dimensional Laplace operator.

In (6) we used (5) and Maxwell's equations for $E_x{}^i$. A similar expression as (6) is obtained for E_y. These equations can now be written in the form

$$\begin{aligned} \nabla_{xy}{}^2 A_x + k^2 A_x &= \mu_0\frac{\partial H_y{}^i}{\partial z} \\ \nabla_{xy}{}^2 A_y + k^2 A_y &= -\mu_0\frac{\partial H_x{}^i}{\partial z} \quad \text{on the disk} \\ \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} &= -\mu_0 H_z{}^i \end{aligned} \tag{7}$$

where $k = \omega(\mu_0\epsilon_0)^{1/2}$ is the free-space wave number.

Eqs. (7) are the basic differential equations which we shall solve for the vector potential $\boldsymbol{A}(x, y)$ on the disk. Knowing $\boldsymbol{A}(x, y)$ we then obtain the surface-current distribution $J(x', y')$ from the integral equation (1).

In order to obtain a unique solution the conditions 1)–4) in Section I have to be fulfilled. Conditions 1) and 2) have been taken care of in (2)–(4). The edge condition 3) will be fulfilled by assuming a suitable current distribution on the disk, and the radiation conditions 4) are secured by (1) for any finite current distribution.

In the following we attempt to find a power-series solution for the current J in terms of (ka), which is expected to converge well for small disks where $(ka) = 2\pi(a/\lambda) < 1$ or $a < \lambda/2\pi$ (λ = free-space wavelength).

First, we consider an expansion of $\boldsymbol{A}$, J and the right-hand side of (7) in powers of (k). We obtain the following expressions:

$$J = J^0 + kJ^1 + k^2J^2 + k^3J^3 + k^4J^4 + \cdots \tag{8}$$

$$\begin{aligned} Je^{-jkr} = {} & J^0 + k(J^1 - jrJ^0) + k^2(J^2 - jrJ^1 - 1/2r^2J^0) \\ & + k^3(J^3 - jrJ^2 - 1/2r^2J^1 + 1/6jr^3J^0) \\ & + k^4(J^4 - jrJ^3 - 1/2r^2J^2 + 1/6jr^3J^1 \\ & + 1/24r^4J^0) + \cdots \end{aligned} \tag{9}$$

$$\boldsymbol{A} = \boldsymbol{A}^0 + k\boldsymbol{A}^1 + k^2\boldsymbol{A}^2 + k^3\boldsymbol{A}^3 + k^4\boldsymbol{A}^4 + \cdots \tag{10}$$

where

$$\begin{aligned} \boldsymbol{A}^0 &= \frac{\mu_0}{4\pi}\int_D J^0\frac{dS}{r} \\ \boldsymbol{A}^1 &= \frac{\mu_0}{4\pi}\left\{\int_D J^1\frac{dS}{r} - j\int_D J^0 dS\right\} \\ \boldsymbol{A}^2 &= \frac{\mu_0}{4\pi}\left\{\int_D J^2\frac{dS}{r} - j\int_D J^1 dS - 1/2\int_D J^0 r dS\right\} \\ \boldsymbol{A}^3 &= \frac{\mu_0}{4\pi}\left\{\int_D J^3\frac{dS}{r} - j\int_D J^2 dS - 1/2\int_D J^1 r dS \right. \\ &\qquad \left. + 1/6j\int_D J^0 r^2 dS\right\} \\ \boldsymbol{A}^4 &= \frac{\mu_0}{4\pi}\left\{\int_D J^4\frac{dS}{r} - j\int_D J^3 dS - 1/2\int_D J^2 r dS \right. \\ &\qquad \left. + 1/6j\int_D J^1 r^2 dS + 1/24j\int_D J^0 r^3 dS\right\}. \end{aligned} \tag{11}$$

For simplification we adopt the following notation:

$$\begin{aligned} \boldsymbol{A}^0 &= \boldsymbol{A}^{00} \\ \boldsymbol{A}^1 &= \boldsymbol{A}^{11} - j\boldsymbol{A}^{01} \\ \boldsymbol{A}^2 &= \boldsymbol{A}^{22} - j\boldsymbol{A}^{12} - 1/2\boldsymbol{A}^{02} \\ \boldsymbol{A}^3 &= \boldsymbol{A}^{33} - j\boldsymbol{A}^{23} - 1/2\boldsymbol{A}^{13} + 1/6j\boldsymbol{A}^{03} \\ \boldsymbol{A}^4 &= \boldsymbol{A}^{44} - j\boldsymbol{A}^{34} - 1/2\boldsymbol{A}^{24} + 1/6j\boldsymbol{A}^{14} + 1/24\boldsymbol{A}^{04} \end{aligned} \tag{12}$$

where the partial vector potentials $\boldsymbol{A}^{mn}$ are defined by (11) and (12).

$$\begin{aligned} \mu_0 \frac{\partial H_y{}^i}{\partial z} &= S^0 + kS^1 + k^2S^2 + k^3S^3 + \cdots \\ -\mu_0 \frac{\partial H_x{}^i}{\partial z} &= T^0 + kT^1 + k^2T^2 + k^3T^3 + \cdots \\ -\mu_0 H_z &= U^0 + kU^1 + k^2U^2 + k^3U^3 + \cdots . \end{aligned} \tag{13}$$

The coefficients S^n, T^n, and U^n in (13) are, of course, functions of the field coordinates x and y and can, therefore, be written as

$$\begin{aligned} S^n &= S_0{}^n + S_1{}^n x + S_2{}^n y + S_3{}^n x^2 + S_4{}^n xy \\ &\quad + S_5{}^n y^2 + \cdots \\ T^n &= T_0{}^n + T_1{}^n y + T_2{}^n x + T_3{}^n y^2 + T_4{}^n yx \\ &\quad + T_5{}^n x^2 + \cdots \\ U^n &= U_0{}^n + U_1{}^n x + U_2{}^n y + U_3{}^n x^2 + U_4{}^n xy \\ &\quad + U_5{}^n y^2 + \cdots . \end{aligned} \tag{14}$$

The coefficients $S_m{}^n$, $T_m{}^n$ and $U_m{}^n$ are constants and can be calculated from the primary field. For the term $k^n x^r y^s$ we obtain

$$\begin{aligned} S_m{}^n &= \frac{1}{n!r!s!} \frac{\partial^{r+s}}{\partial x^r \partial y^s} \frac{\partial^n}{\partial k^n} \left[\mu_0 \left(\frac{\partial H_y{}^i}{\partial z} \right) \right]_{\substack{x=y=z=0 \\ k=0}} \\ T_m{}^n &= \frac{1}{n!r!s!} \frac{\partial^{r+s}}{\partial x^r \partial y^s} \frac{\partial^n}{\partial k^n} \left[-\mu_0 \left(\frac{\partial H_y{}^i}{\partial z} \right) \right]_{\substack{x=y=z=0 \\ k=0}} \\ U_m{}^n &= \frac{1}{n!r!s!} \frac{\partial^{r+s}}{\partial x^r \partial y^s} \frac{\partial^n}{\partial_k n} \left[-\mu_0 H_z{}^i \right]_{\substack{x=y=z=0. \\ k=0}} \end{aligned} \tag{15}$$

Where all derivatives are taken for $x=y=z=k=0$. The subscript m can be found by inspection of (14) or from the relation

$$m = \left(\sum_{t=1}^{r+s} t \right) + s. \tag{15a}$$

We now substitute (10) and (13) in the basic equation (7). Equating equal powers of k we obtain for the nth-order approximation, *i.e.*, for k^n, the following relations:

$$\begin{aligned} \nabla^2 A_x{}^n + A_x{}^{n-2} &= S^n \\ \nabla^2 A_y{}^n + A_y{}^{n-2} &= T^n \\ \frac{\partial A_y{}^n}{\partial x} - \frac{\partial A_x{}^n}{\partial y} &= U^n. \end{aligned} \tag{16}$$

It is interesting to note that $\boldsymbol{A}^n$ does depend on the value of $\boldsymbol{A}^{n-2}$. Consider now (16) in terms of the partial vector potentials as defined in (11) and (12). It is readily seen that the second term in these expressions, $\boldsymbol{A}^{(n-1)n}$, is independent of the field coordinates x and y, so that all its derivatives vanish identically. Summarizing we find that in the calculation for $\boldsymbol{A}^n$ the expressions for $\boldsymbol{A}^{n-1}$ and $\boldsymbol{A}^{(n-1)n}$ or, in terms of the surface currents, the term $\boldsymbol{J}^{n-1}$, do not appear. From this follows that the zero- and first-order approximation for $\boldsymbol{J}$ are not related to each other and can be evaluated from the expansion coefficients of the primary field only. For higher-order approximations, however, we have to use the results of the previous calculations. This can be clearly seen, if we rewrite (16) and keep only the term $\boldsymbol{A}^{nn}$ on the left-hand side. We then obtain

$$\nabla^2 A_x{}^{nn} = V^n \tag{17a}$$

$$\nabla^2 A_y{}^{nn} = W^n \tag{17b}$$

$$\frac{\partial A_y{}^{nn}}{\partial x} - \frac{\partial A_x{}^{nn}}{\partial y} = X^n \tag{17c}$$

where V^n, W^n and X^n are functions of the field coordinates x and y and can be written as follows:

$$\begin{aligned} V^n &= V_0{}^n + V_1{}^n x + V_2{}^n y + V_3{}^n x^2 + V_4{}^n xy \\ &\quad + V_5{}^n y^2 + \cdots = \sum_{r,s} V_m{}^n x^r y^s \\ W^n &= W_0{}^n + W_1{}^n y + W_2{}^n x + W_3{}^n y^2 + W_4{}^n yx \\ &\quad + W_5{}^n x^2 + \cdots = \sum_{r,s} W_m{}^n y^r x^s \\ X^n &= X_0{}^n + X_1{}^n x + X_2{}^n y + X_3{}^n x^2 + X_4{}^n xy \\ &\quad + X_5{}^n y^2 + \cdots = \sum_{r,s} X_m{}^n x^r y^s. \end{aligned} \tag{18}$$

Again m is found from (15a).

$V_m{}^n$, $W_m{}^n$ and $X_m{}^n$ can be calculated from the coefficients $S_m{}^n$, $T_m{}^n$ and $U_m{}^n$ for the primary field and from the results of the $(n-2)$ and lower-order approximations as will be shown explicitly in (29). The $A_x{}^{nn}$ and $A_y{}^{nn}$ can then be obtained from (17) by straightforward integration in form of power series in x and y. The main problem is now to find a solution for the integral equation for the current $\boldsymbol{J}^n$. This equation is obtained from (11) and (12):

$$\boldsymbol{A}^{nn}(x, y) = \frac{\mu_0}{4\pi} \int_D \boldsymbol{J}^n(x', y') \frac{dx'dy'}{[(x-x')^2 + (y-y')^2]^{1/2}}. \tag{19}$$

Formal Solution for the Surface Current Distribution

The kind of integral which appears in (19) has been investigated by Bouwkamp [45], [49], [50]. He found

that if we write the current distribution in the following form:

$$\frac{\mu_0}{4\pi} J_x(x', y') = \frac{f_x(x', y')}{\pi^2(a^2 - \rho'^2)^{1/2}}$$

$$\frac{\mu_0}{4\pi} J_y(x', y') = \frac{f_y(x', y')}{\pi^2(a^2 - \rho'^2)^{1/2}}$$

$$(\rho')^2 = (x')^2 + (y')^2 \tag{20}$$

where f_x and f_y are polynomials in x' and y', the integral in (19) is also a polynomial in x and y, and is of the same order as $f(x', y')$. Table I in the Appendix gives a solution for the following integral:

$$G(x, y) = \frac{1}{\pi^2} \int_D \frac{f(x', y')dS}{(a^2 - \rho'^2)^{1/2}[(x - x')^2 + (y - y')^2]^{1/2}}. \tag{21}$$

The integration is over the surface D of the disk; $f(x', y')$ is a polynomial in x' and y'. Only terms up to the fourth power in x' and y' have been calculated. Computations of higher-order terms are possible yet they become, in principle, extremely tedious. It is noteworthy that $G(x, y)$ is also a polynomial of the same order as $f(x', y')$.

In order to obtain a unique solution, we have to ascertain that the edge conditions are fulfilled. It can be shown that they are equivalent to the requirement that the current component normal to the edge vanishes at the rim of the disk. The radial current is

$$J_r = J_x \cos\varphi' + J_y \sin\varphi' = (1/\rho')(x'J_x + y'J_y)$$

and thus we must have

$$\frac{4}{\pi\mu_0} \lim_{\rho' \to a} \frac{x'f_x + y'f_y}{\rho'(a^2 - \rho'^2)^{1/2}} = 0. \tag{22}$$

We further require that the total charge on the disk remains finite. The charge density σ is given by

$$\sigma = \frac{j}{\omega} \nabla \cdot \boldsymbol{J} = \frac{j}{\omega} \frac{4}{\pi\mu_0} \nabla \cdot \frac{\boldsymbol{f}}{(a^2 - \rho'^2)^{1/2}}$$

$$= \frac{4j}{\omega\pi\mu_0} \left[\left(\frac{\partial f_x}{\partial x'} + \frac{\partial f_y}{\partial y'} \right) \frac{1}{(a^2 - \rho'^2)^{1/2}} + \frac{x'f_x + y'f_x}{(a^2 - \rho'^2)^{3/2}} \right]. \tag{23}$$

The last term in (23) is only integrable over the disk if

$$(x'f_x + y'f_y) = D(x', y')(a^2 - \rho'^2) \tag{24}$$

where $D(x', y')$ is also a polynomial in x' and y' as will be shown in the following calculations. This condition, however, is sufficient to satisfy (22) and is, therefore, equivalent to the edge condition.

The current-distribution function is now expanded in terms of k as follows:

$$\boldsymbol{f} = \boldsymbol{f}^0 + k\boldsymbol{f}^1 + k^2\boldsymbol{f}^2 + k^3\boldsymbol{f}^3 + \cdots. \tag{25}$$

We then write for the polynomial

$$\boldsymbol{f}^n(x', y') = f_x^n(x', y')\boldsymbol{a}_x + f_y^n(x', y')\boldsymbol{a}_y$$

$$f_x^n(x', y') = a_0^n + a_1^n x' + a_2^n y' + a_3^n x'^2 + a_4^n x'y' + a_5^n y'^2 + a_6^n x'^3 + a_7^n x'^2 y' + a_8^n x'y'^2 + a_9^n y'^3 + a_{10}^n x'^4 + a_{11}^n x'^3 y' + a_{12}^n x'^2 y'^2 + a_{13}^n x'y'^3 + a_{14}^n y'^4 + \cdots = \sum_{r,s} a_m^n x^r y^s \tag{26a}$$

$$f_y^n(x', y') = b_0^n + b_1^n y' + b_2^n x' + b_3^n y'^2 + b_4^n y'x' + b_5^n x'^2 + b_6^n y'^3 + b_7^n y'^2 x' + b_8^n y'x'^2 + b_9^n x'^3 + b_{10}^n y'^4 + b_{11}^n y'^3 x' + b_{12}^n y'^2 x'^2 + b_{13}^n y'x'^3 + b_{14}^n x'^4 + \cdots = \sum_{r,s} b_m^n y^r x^s. \tag{26b}$$

The edge conditions (24) become:

$$(x'f_x^n + y'f_y^n) = (d_0^n + d_1^n x' + d_2^n y' + d_3^n x'^2 + d_4^n x'y' + d_5^n y'^2 + d_6^n x'^3 + d_7^n x'^2 y' + d_8^n x'y'^2 + d_9^n y'^3)(a^2 - x'^2 - y'^2). \tag{27}$$

Calculations of the Current Coefficients a and b

Eqs. (17)–(20), (26), and (27) determine now a total of 40 coefficients a^n and b^n, if in $f(x'y')$ terms up to the fourth power in x' and y' are considered. Detailed calculations are presented in a separate report [52]. Here a short outline of the procedure is given: the coefficients a_m^n and b_m^n are found in terms of V_m^n, W_m^n, and X_m^n defined in (17) and (18)

$$a_m^n = a_m^n(V_{m'}^n, W_{m'}^n, X_{m'}^n)$$

$$b_m^n = b_m^n(V_{m'}^n, W_{m'}^n, X_{m'}^n). \tag{28}$$

The prime on the subscripts m in the brackets indicate that terms with a different subscript than m are involved.

Our next problem consists in expressing V_m^n, W_m^n, and X_m^n in terms of the field quantities S_m^n, T_m^n, and U_m^n. If we use the partial vector potentials given in (11), (12), and (16) and keep only the terms with A^n on the left-hand side, we obtain together with (17), (20) and (26)

$$V_m^n = V_m^n(S_m^n, a_{m'}^{n-2}, a_{m'}^{n-3}, \cdots)$$

$$W_m^n = W_m^n(T_m^n, b_{m'}^{n-2}, b_{m'}^{n-3}, \cdots)$$

$$X_m^n = X_m^n(U_m^n, a_{m'}^{n-2}, a_{m'}^{n-3}, \cdots, b_{m'}^{n-2}, b_{m'}^{n-3}, \cdots) \tag{29}$$

where only current coefficients of $(n-2)$ and lower order are involved. Putting these expressions back in (28) and replacing all the current coefficients of lower order by an iterative process we finally obtain an expression which involves only the field coefficients S_m^n, T_m^n, and U_m^n

$$a_m^n = a_m^n(S_{m'}^{n'}, T_{m'}^{n'}, U_{m'}^{n'})$$

$$b_m^n = b_m^n(S_{m'}^{n'}, T_{m'}^{n'}, U_{m'}^{n'}). \tag{30}$$

Following (25) we define the total current coefficients by

$$a_m = a_m^0 + ka_m^1 + k^2a_m^2 + k^3a_m^3 + \cdots$$
$$b_m = b_m^0 + kb_m^1 + k^2b_m^2 + k^3b_m^3 + \cdots \qquad (31)$$

and similarly for the expansion coefficients of the incident field

$$S_m = \sum_{n=0}^{\infty} k^n S_m^n = \sum_{n=0}^{\infty} \left\{ \frac{k^n}{n!} \frac{\partial^n}{\partial k^n} \left[\frac{\partial^{r+s}}{\partial x^r \partial y^s} \cdot \left(\mu_0 \frac{\partial H_y^i}{\partial z} \right) \right] \right\}_{x=y=z=k=0} \frac{1}{r!s!}$$
$$= \frac{1}{r!s!} \left[\frac{\partial^{r+s}}{\partial_x r \partial_y s} \left(\mu_0 \frac{\partial H_y^i}{\partial z} \right) \right]_{x=y=z=0}$$
$$T_m = \sum_{n=0}^{\infty} k^n T_m^n = \sum_{n=0}^{\infty} \left\{ \frac{k^n}{n!} \frac{\partial^n}{\partial k^n} \left[\frac{\partial^{r+s}}{\partial x^r \partial y^s} \cdot \left(-\mu_0 \frac{\partial H_x^i}{\partial z} \right) \right] \right\}_{x=y=z=k=0} \frac{1}{r!s!}$$
$$= \frac{1}{r!s!} \left[\frac{\partial^{r+s}}{\partial y^r \partial x^s} \left(-\mu_0 \frac{\partial H_x^i}{\partial z} \right) \right]_{x=y=z=0}$$
$$U_m = \sum_{n=0}^{\infty} k^n U_m^n = \sum_{n=0}^{\infty} \left\{ \frac{k^n}{n!} \frac{\partial^n}{\partial k^n} \left[\frac{\partial^{r+s}}{\partial y^r \partial x^s} \cdot (-\mu_0 H_z^i) \right] \right\}_{x=y=z=k=0} \frac{1}{r!s!}$$
$$= \frac{1}{r!s!} \left[\frac{\partial^{r+s}}{\partial y^r \partial x^s} (-\mu_0 Hz^i) \right]_{x=y=z=0}. \qquad (32)$$

The right-hand side is obtained by using (15) and Taylor's expansion theorem.

It turns out that all current coefficients belonging to different powers of k^n but having the same subscript m have the same functional dependence on S_m^n, T_m^n, and U_m^n. This allows us to write the total current coefficients in terms of the primary field directly by using (32). After some calculations we finally obtain

$$a_0 = \frac{a^2}{315}\mu_0 \left\{ 210 \left(-2 \frac{\partial H_y}{\partial z} + \frac{\partial H_z}{\partial y} \right) + a^2 \left(-28 \frac{\partial^3 H_y}{\partial x^2 \partial z} - 28 \frac{\partial^3 H_y}{\partial y^2 \partial z} + 21 \frac{\partial^3 H_z}{\partial y^3} + 21 \frac{\partial^3 H_z}{\partial x^2 \partial y} \right) + (ka)^2 \left(-196 \frac{\partial H_y}{\partial z} + 168 \frac{\partial H_z}{\partial y} \right) \right\} + j \frac{8a^2}{9\pi} \mu_0 (ka)^3 \left(\frac{\partial H_y}{\partial z} - \frac{\partial H_z}{\partial y} \right)$$

$$a_1 = \frac{a^2}{30}\mu_0 \left\{ -22 \frac{\partial^2 H_y}{\partial x \partial z} - 2 \frac{\partial^2 H_x}{\partial y \partial z} + 8 \frac{\partial^2 H_z}{\partial x \partial y} \right\}$$

$$a_2 = \frac{\mu_0}{30} \left\{ 30 H_z + a^2 \left(-13 \frac{\partial^2 H_y}{\partial y \partial z} + 11 \frac{\partial^2 H_x}{\partial x \partial z} - 8 \frac{\partial^2 H_z}{\partial x^2} \right) - (ka)^2 H_z \right\} + j \frac{4a^2}{9\pi} \mu_0 (ka)^3 H_z$$

$$a_3 = \frac{\mu_0}{315} \left\{ 210 \left(2 \frac{\partial H_y}{\partial z} - \frac{\partial H_z}{\partial y} \right) + a^2 \left(-52 \frac{\partial^3 H_y}{\partial x^2 \partial z} + 44 \frac{\partial^3 H_y}{\partial y^2 \partial z} - 9 \frac{\partial^3 H_z}{\partial x^2 \partial y} - 33 \frac{\partial^3 H_z}{\partial y^3} \right) + (ka)^2 \left(224 \frac{\partial H_y}{\partial z} - 195 \frac{\partial H_z}{\partial y} \right) \right\} + j \frac{8\mu_0}{9\pi} (ka)^3 \left(-\frac{\partial H_y}{\partial z} + \frac{\partial H_z}{\partial y} \right)$$

$$a_4 = \frac{\mu_0}{315} \left\{ 210 \left(-\frac{\partial H_x}{\partial z} + 2 \frac{\partial H_z}{\partial x} \right) + a^2 \left(-27 \frac{\partial^3 H_x}{\partial y^2 \partial z} + 165 \frac{\partial^3 H_x}{\partial x^2 \partial z} - 150 \frac{\partial^3 H_z}{\partial x^3} - 102 \frac{\partial^3 H_z}{\partial x \partial y^2} \right) + (ka)^2 \left(-168 \frac{\partial H_x}{\partial z} - 12 \frac{\partial H_z}{\partial x} \right) \right\} + j \frac{8\mu_0}{9\pi} (ka)^3 \left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} \right)$$

$$a_5 = \frac{\mu_0}{315} \left\{ 210 \left(\frac{\partial H_y}{\partial z} + \frac{\partial H_z}{\partial y} \right) + a^2 \left(17 \frac{\partial^3 H_y}{\partial x^2 \partial z} - 79 \frac{\partial^3 H_y}{\partial y^2 \partial z} + 9 \frac{\partial^3 H_z}{\partial x^2 \partial y} + 33 \frac{\partial^3 H_z}{\partial y^3} \right) + (ka)^2 \left(56 \frac{\partial H_y}{\partial z} - 15 \frac{\partial H_z}{\partial y} \right) \right\}$$

$$a_6 = \frac{\mu_0}{30} \left\{ 22 \frac{\partial^2 H_y}{\partial x \partial z} + 2 \frac{\partial^2 H_x}{\partial y \partial z} - 8 \frac{\partial^2 H_z}{\partial x \partial y} \right\}$$

$$a_7 = \frac{\mu_0}{30} \left\{ 14 \frac{\partial^2 H_y}{\partial y \partial z} - 18 \frac{\partial^2 H_x}{\partial x \partial z} + 24 \frac{\partial^2 H_z}{\partial x^2} + 7k^2 H_z \right\}$$

$$a_8 = \frac{\mu_0}{30} \left\{ 14 \frac{\partial^2 H_y}{\partial x \partial z} - 6 \frac{\partial^2 H_x}{\partial y \partial z} + 24 \frac{\partial^2 H_z}{\partial x \partial y} \right\}$$

$$a_9 = \frac{\mu_0}{30} \left\{ 22 \frac{\partial^2 H_y}{\partial y \partial z} + 6 \frac{\partial^2 H_x}{\partial x \partial z} - 8 \frac{\partial^2 H_z}{\partial x^2} - 9k^2 H_z \right\}$$

$$a_{10} = \frac{\mu_0}{315} \left\{ 80 \frac{\partial^3 H_y}{\partial x^2 \partial z} - 16 \frac{\partial^3 H_y}{\partial y^2 \partial z} - 12 \frac{\partial^3 H_z}{\partial x^2 \partial y} + 12 \frac{\partial^3 H_z}{\partial y^3} + k^2 \left(-28 \frac{\partial H_y}{\partial z} + 27 \frac{\partial H_z}{\partial y} \right) \right\}$$

$$a_{11} = \frac{\mu_0}{315} \left\{ 12 \frac{\partial^3 H_x}{\partial y^2 \partial z} - 204 \frac{\partial^3 H_x}{\partial x^2 \partial z} + 216 \frac{\partial^3 H_z}{\partial x^3} + 120 \frac{\partial^3 H_z}{\partial x \partial y^2} + k^2 \left(21 \frac{\partial H_x}{\partial z} + 150 \frac{\partial H_z}{\partial x} \right) \right\}$$

$$a_{12}=\frac{\mu_0}{315}\left\{52\frac{\partial^3 H_y}{\partial x^2\partial y}+124\frac{\partial^3 H_y}{\partial y^2\partial z}+72\frac{\partial^3 H_z}{\partial x^2\partial y}-72\frac{\partial^3 H_z}{\partial y^3}+k^2\left(-35\frac{\partial H_y}{\partial z}-36\frac{\partial H_z}{\partial y}\right)\right\}$$

$$a_{13}=\frac{\mu_0}{315}\left\{-12\frac{\partial^3 H_x}{\partial y^2\partial z}-132\frac{\partial^3 H_x}{\partial x^2\partial z}+120\frac{\partial^3 H_z}{\partial x^3}+216\frac{\partial^3 H_z}{\partial x\partial y^2}+k^2\left(21\frac{\partial H_x}{\partial z}+102\frac{\partial H_z}{\partial x}\right)\right\}$$

$$a_{14}=\frac{\mu_0}{315}\left\{-4\frac{\partial^3 H_y}{\partial x^2\partial z}+68\frac{\partial^3 H_y}{\partial y^2\partial z}-12\frac{\partial^3 H_z}{\partial x^2\partial y}+12\frac{\partial^3 H_z}{\partial y^3}+k^2\left(-7\frac{\partial H_y}{\partial z}-15\frac{\partial H_z}{\partial y}\right)\right\}. \tag{33}$$

$$b_0=\frac{a^2}{315}\mu_0\left\{210\left(2\frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}\right)+a^2\left(28\frac{\partial^3 H_x}{\partial y^2\partial z}+28\frac{\partial^3 H_x}{\partial x^2\partial z}-21\frac{\partial^3 H_z}{\partial x^3}-21\frac{\partial^3 H_z}{\partial x\partial y^2}\right)+(ka)^2\left(196\frac{\partial H_x}{\partial z}-168\frac{\partial H_z}{\partial x}\right)\right\}+j\frac{8a^2}{9\pi}\mu_0(ka)^3\left(-\frac{\partial H_x}{\partial z}+\frac{\partial H_z}{\partial x}\right)$$

$$b_1=\frac{a^2}{30}\mu_0\left\{22\frac{\partial^2 H_x}{\partial y\partial z}+2\frac{\partial^2 H_y}{\partial x\partial z}-8\frac{\partial^2 H_z}{\partial x\partial y}\right\}$$

$$b_2=\frac{\mu_0}{30}\left\{-30H_z+a^2\left(13\frac{\partial^2 H_x}{\partial x\partial z}+11\frac{\partial^2 H_y}{\partial y\partial z}+8\frac{\partial^2 H_z}{\partial y^2}\right)+9(ka)^2H_z\right\}-j\frac{4a^2}{9\pi}\mu_0(ka)^3H_z$$

$$b_3=\frac{\mu_0}{315}\left\{210\left(-2\frac{\partial H_x}{\partial z}+\frac{\partial H_z}{\partial x}\right)+a^2\left(52\frac{\partial^3 H_x}{\partial y^2\partial z}-44\frac{\partial^3 H_x}{\partial x^2\partial z}+9\frac{\partial^3 H_z}{\partial x\partial y^2}+33\frac{\partial^3 H_z}{\partial x^3}\right)+(ka)^2\left(-224\frac{\partial H_x}{\partial z}+195\frac{\partial H_z}{\partial x}\right)\right\}+j\frac{8\mu_0}{9\pi}(ka)^3\left(\frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}\right)$$

$$b_4=\frac{\mu_0}{315}\left\{210\left(\frac{\partial H_y}{\partial z}-2\frac{\partial H_z}{\partial y}\right)+a^2\left(27\frac{\partial^3 H_y}{\partial x^2\partial z}-165\frac{\partial^3 H_y}{\partial y^2\partial z}+150\frac{\partial^3 H_z}{\partial y^3}+102\frac{\partial^3 H_z}{\partial x^2\partial y}\right)+(ka)^2\left(168\frac{\partial H_y}{\partial z}+12\frac{\partial H_z}{\partial y}\right)\right\}+j\frac{8}{9\pi}\mu_0(ka)^3\left(-\frac{\partial H_y}{\partial z}+\frac{\partial H_z}{\partial y}\right)$$

$$b_5=\frac{\mu_0}{315}\left\{210\left(-\frac{\partial H_x}{\partial z}-\frac{\partial H_z}{\partial x}\right)+a^2\left(-17\frac{\partial^3 H_y}{\partial y^2\partial z}+79\frac{\partial^3 H_x}{\partial x^2\partial z}-9\frac{\partial^3 H_z}{\partial x\partial y^2}-33\frac{\partial^3 H_z}{\partial x^3}\right)+(ka)^2\left(-56\frac{\partial H_x}{\partial z}+15\frac{\partial H_z}{\partial x}\right)\right\}$$

$$b_6=\frac{\mu_0}{30}\left\{-22\frac{\partial^2 H_x}{\partial y\partial z}-2\frac{\partial^2 H_y}{\partial x\partial z}+8\frac{\partial^2 H_z}{\partial x\partial y}\right\}$$

$$b_7=\frac{\mu_0}{30}\left\{-14\frac{\partial^2 H_x}{\partial x\partial z}+18\frac{\partial^2 H_y}{\partial y\partial z}-24\frac{\partial^2 H_z}{\partial y^2}-7k^2H_z\right\}$$

$$b_8=\frac{\mu_0}{30}\left\{-14\frac{\partial^2 H_x}{\partial y\partial z}+6\frac{\partial^2 H_y}{\partial x\partial z}-24\frac{\partial^2 H_z}{\partial x\partial y}\right\}$$

$$b_9=\frac{\mu_0}{30}\left\{-22\frac{\partial^2 H_x}{\partial x\partial z}-6\frac{\partial^2 H_y}{\partial y\partial z}+8\frac{\partial^2 H_z}{\partial y^2}+9k^2H_z\right\}$$

$$b_{10}=\frac{\mu_0}{315}\left\{-80\frac{\partial^3 H_x}{\partial y^2\partial z}+16\frac{\partial^3 H_x}{\partial x^2\partial z}+12\frac{\partial^3 H_z}{\partial x\partial y^2}-12\frac{\partial^3 H_z}{\partial x^3}+k^2\left(28\frac{\partial H_x}{\partial z}-27\frac{\partial H_z}{\partial x}\right)\right\}$$

$$b_{11}=\frac{\mu_0}{315}\left\{-12\frac{\partial^3 H_y}{\partial x^2\partial z}+204\frac{\partial^3 H_y}{\partial y^2\partial z}-216\frac{\partial^3 H_z}{\partial y^3}-120\frac{\partial^3 H_z}{\partial x^2\partial y}+k^2\left(-21\frac{\partial H_y}{\partial z}-150\frac{\partial H_z}{\partial y}\right)\right\}$$

$$b_{12}=\frac{\mu_0}{315}\left\{-52\frac{\partial^3 H_x}{\partial y^2\partial z}-124\frac{\partial^3 H_x}{\partial x^2\partial z}-72\frac{\partial^3 H_z}{\partial x\partial y^2}+72\frac{\partial^3 H_z}{\partial x^3}+k^2\left(35\frac{\partial H_x}{\partial z}+36\frac{\partial H_z}{\partial x}\right)\right\}$$

$$b_{13}=\frac{\mu_0}{315}\left\{12\frac{\partial^3 H_y}{\partial x^2\partial z}+132\frac{\partial^3 H_y}{\partial y^2\partial z}-120\frac{\partial^3 H_z}{\partial y^3}-216\frac{\partial^3 H_z}{\partial x^2\partial y}+k^2\left(-21\frac{\partial H_y}{\partial z}-102\frac{\partial H_z}{\partial y}\right)\right\}$$

$$b_{14}=\frac{\mu_0}{315}\left\{4\frac{\partial^3 H_x}{\partial y^2\partial z}-68\frac{\partial^3 H_x}{\partial x^2\partial z}+12\frac{\partial^3 H_z}{\partial x\partial y^2}-12\frac{\partial^3 H_z}{\partial x^3}+k^2\left(7\frac{\partial H_x}{\partial z}+15\frac{\partial H_z}{\partial x}\right)\right\}. \tag{34}$$

The superscript i for the incident field has been omitted. All magnetic field components and their derivatives should, of course, be evaluated at the center of the disk.

Eqs. (33) and (34) represent the solution of our problem from which we shall derive the results in the following sections.

III. Induced Electric-Dipole Moment

The induced electric-dipole moment is defined by

$$\boldsymbol{P} = \int_D \boldsymbol{p}\sigma(\boldsymbol{p})dS \tag{35}$$

where

$$\sigma = \frac{j}{\omega}\nabla\cdot\boldsymbol{J} \tag{36}$$

is the electric-surface-charge density and $\boldsymbol{p}$ the radius vector on the disk.

Obviously only odd-power terms in x' and y' for $\sigma(x', y')$ (or even-power terms for $J(x', y')$) contribute to $\boldsymbol{P}$.

Using the results in Section II and Maxwell's equation we obtain

$$P_x = \frac{16}{3}a^3\epsilon_0\left[E_x^i + \frac{(ka)^2}{30}\left(13E_x^i - \frac{3}{k^2}\frac{\partial^2 E_x^i}{\partial z^2} + \frac{2j}{\omega\epsilon_0}\frac{\partial H_z^i}{\partial y}\right) - \frac{8j}{9\pi}(ka)^3E_x^i\right]_0$$

$$P_y = \frac{16}{3}a^3\epsilon_0\left[E_y^i + \frac{(ka)^2}{30}\left(13E_y^i - \frac{3}{k^2}\frac{\partial^2 E_y^i}{\partial z^2} - \frac{2j}{\omega\epsilon_0}\frac{\partial H_z^i}{\partial x}\right) - \frac{8j}{9\pi}(ka)^3E_y^i\right]_0. \tag{37}$$

The subscript 0 at the right-hand side of the main bracket in (37) indicates that all terms have to be evaluated for $x=y=z=0$.

The first terms in (37) represent the well-known first-order approximations calculated by Bethe [21].

Eq. (37) shows also that in the first-order approximation $\boldsymbol{P}$ is in the direction of the tangential electric field at the center of the disk. The higher-order terms, however, lead to cross polarization, due to the terms $\partial H_z^i/\partial y$ and $\partial H_z^i/\partial x$.

For the case of an incident plane wave the electric-dipole moment depends also on the angle of incidence θ_i and is given by (see Fig. 1)

$$P_x = \frac{16}{3}a^3\epsilon_0\left[1 + \left(\frac{8}{15} - \frac{1}{10}\sin^2\theta_i\right)(ka)^2 - j\frac{8}{9\pi}(ka)^3\right]E_x^i(0,0,0)$$

$$P_y = \frac{16}{3}a^3\epsilon_0\left[1 + \left(\frac{8}{15} - \frac{1}{6}\sin^2\theta_i\right)(ka)^2 - j\frac{8}{9\pi}(ka)^3\right]E_y^i(0,0,0). \tag{38}$$

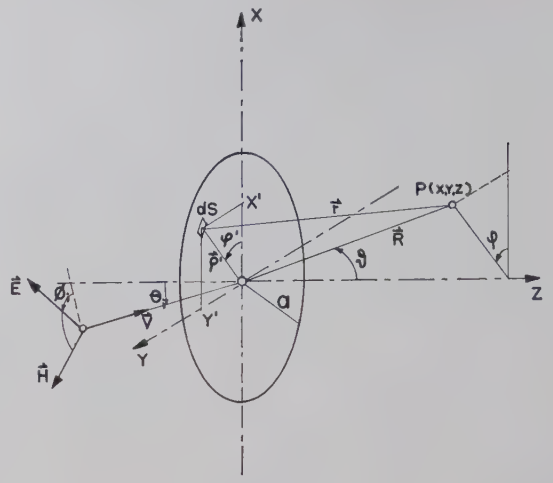


Fig. 1.

The case for normal incidence has been treated by Bouwkamp [23] up to the fifth-order approximation. From his current distribution, one obtains

$$\boldsymbol{P} = \frac{16}{3}a^3\epsilon_0\left[1 + \frac{8}{15}(ka)^2 - j\frac{8}{9\pi}(ka)^3 + \frac{16}{105}(ka)^4 - j\frac{176}{225\pi}(ka)^5\right]\boldsymbol{E}(0,0,0). \tag{39}$$

IV. Induced Magnetic-Dipole Moment

The magnetic-dipole moment is defined by

$$\boldsymbol{M} = \frac{1}{2}\int_D \boldsymbol{p}\times\boldsymbol{j}(\boldsymbol{p})dS. \tag{40}$$

In the case of a plane disk $\boldsymbol{M}$ has only a component M_z along the disk axis. It is easily seen that only odd power terms in $J(x', y')$ contribute.

Again we can express M_z in terms of the primary field

$$M_z = -\frac{8a^3}{3}\left\{H_z^i - \frac{1}{10}(ka)^2\left[3H_z^i + \frac{1}{k^2}\frac{\partial^2 H_z^i}{\partial z^2}\right] + j\frac{4}{9\pi}(ka)^3H_z^i\right\}_0. \tag{41}$$

The bracketed term is evaluated at $x=y=z=0$.

For a plane wave incident at an angle θ_i (see Fig. 1), (41) becomes

$$M_z = -\frac{8}{3}a^3\left\{1 - \frac{1}{10}(2+\sin^2\theta_i)(ka)^2 + j\frac{4}{9\pi}(ka)^3\right\}H_z^i(0,0,0). \tag{42}$$

For normal incidence the induced magnetic-dipole moment vanishes.

V. Far-Zone Fields

Given the current distribution we can find the scattered field in a straightforward manner from the vector potential as given in (1). Unfortunately a general solution of this integral is not available. It is, however, possible to evaluate it for field points which are at distances large compared with the disk radius. In this case we can write

$$\boldsymbol{A} = \frac{\mu_0}{4\pi}\int_D J\frac{e^{-jkr}}{r}\,dS' \xrightarrow[R\to\infty]{} \frac{\mu_0}{4\pi}\frac{e^{-jkR}}{R} \cdot \int_D J\exp\left[jk\rho'\sin\vartheta\cos(\varphi-\varphi')\right]dS' \tag{43}$$

where R, ϑ, φ are the spherical coordinates of the field point, ρ' and φ' are the integration variables as shown in Fig. 1. Expanding the exponential in powers of k up to the term k^3 and using (8) one obtains an expression for $\boldsymbol{A}$ correct to the third order in k. We express the field in spherical coordinates

$$\begin{aligned} A_r &= A_x\cos\varphi\sin\vartheta + A_y\sin\varphi\sin\vartheta \\ A_\vartheta &= A_x\cos\varphi\cos\vartheta + A_y\sin\varphi\cos\vartheta \\ A_\phi &= -A_x\sin\varphi + A_y\cos\varphi. \end{aligned} \tag{44}$$

It turns out that for large distances the field components become

$$\begin{aligned} H_\vartheta &= -\left(\frac{\epsilon_0}{\mu_0}\right)^{1/2} E_\varphi = \frac{jk}{\mu_0}A_\varphi \\ H_\varphi &= \left(\frac{\epsilon_0}{\mu_0}\right)^{1/2} E_\vartheta = -\frac{jk}{\mu_0}A_\vartheta. \end{aligned} \tag{45}$$

After performing the straightforward but tedious computations one finally obtains in terms of the incident field

$$\begin{aligned} H_\vartheta = \frac{e^{-jkR}}{\pi R}(ka)^2 a\Bigg\{ & \frac{1}{4(\mu_0\epsilon_0)^{1/2}a^3}\left[-P_y\cos\varphi + P_x\sin\varphi\right] \\ & + \sin\vartheta\Bigg[-\frac{1}{4a^3}M_z - \frac{4}{45}(ka)^2 H_z{}^i \\ & + j\frac{8}{45\omega\mu_0}(ka)^2\Bigg[-\frac{\partial E_x{}^i}{\partial y}\cos^2\varphi + \frac{\partial E_y{}^i}{\partial x}\sin^2\varphi \\ & + \left(\frac{\partial E_x{}^i}{\partial x} - \frac{\partial E_y{}^i}{\partial y}\right)\sin\varphi\cos\varphi\Bigg]\Bigg] \\ & + \frac{2}{45}(ka)^2\sin^2\vartheta\Bigg[\left(5\sqrt{\frac{\epsilon_0}{\mu_0}}E_y{}^i + j\frac{2}{k}\frac{\partial H_z{}^i}{\partial x}\right)\cos\varphi \\ & + \left(-5\sqrt{\frac{\epsilon_0}{\mu_0}}E_x{}^i + j\frac{2}{k}\frac{\partial H_z{}^i}{\partial y}\right)\sin\varphi\Bigg] \\ & - \frac{1}{15}\sin^3\vartheta(ka)^2 H_z{}^i\Bigg\}_0 \end{aligned} \tag{46a}$$

$$\begin{aligned} H_\varphi = \frac{e^{-jkR}}{\pi R}(ka)^2 a\Bigg\{ & \frac{1}{4(\mu_0\epsilon_0)^{1/2}a^3}\left[P_x\cos\varphi + P_y\sin\varphi\right] \\ & + \frac{(ka)^2}{45}\Bigg[\frac{j}{\omega\mu_0}\sin\vartheta\Bigg[-6\frac{\partial E_z{}^i}{\partial z} - 8\frac{\partial E_y{}^i}{\partial y}\cos^2\varphi \\ & - 8\frac{\partial E_x{}^i}{\partial x}\sin^2\varphi + 8\left(\frac{\partial E_x{}^i}{\partial y} + \frac{\partial E_y{}^i}{\partial x}\right)\sin\varphi\cos\varphi\Bigg] \\ & + 6\sqrt{\frac{\epsilon_0}{\mu_0}}\sin^2\vartheta\left[-E_x{}^i\cos\varphi - E_y{}^i\sin\varphi\right]\Bigg]\Bigg\}_0\cos\vartheta. \end{aligned} \tag{46b}$$

Again all the fields are evaluated for $x=y=z=0$.

For a plane wave at oblique incidence we obtain

$$\begin{aligned} H_\vartheta = \frac{e^{-jkR}}{\pi R}(ka)^2 a H_0\Bigg\{ & \frac{2}{3}\sin\phi_i\sin\theta_i\sin\vartheta \\ & + \frac{4}{3}\cos\phi_i\cos\theta_i\sin\varphi - \frac{4}{3}\sin\phi_i\cos\varphi \\ & + (ka)^2\frac{1}{45}\Bigg[(32-6\sin^2\theta_i)\cos\phi_i\cos\theta_i\sin\varphi \\ & - (32-10\sin^2\theta_i)\sin\phi_i\cos\varphi \\ & + \sin\Big(8\cos\phi_i\cos\theta_i\sin\theta_i\sin\varphi\cos\varphi \\ & + -(2-3\sin^2\theta_i - 8\cos^2\varphi)\sin\phi_i\sin\theta_i\Big) \\ & + \sin^2\vartheta\Big(-10\cos\phi_i\cos\theta_i\sin\varphi \\ & + (10+4\sin^2\theta_i)\sin\phi_i\cos\varphi\Big) \\ & - 3\sin^3\vartheta\sin\phi_i\sin\theta_i\Bigg]\Bigg\} \end{aligned} \tag{47a}$$

$$\begin{aligned} H_\varphi = \frac{e^{-jkR}}{\pi R}(ka)^2 a H_0\cos\vartheta\Bigg\{ & \frac{4}{3}(\cos\phi_i\cos\theta_i\cos\varphi \\ & + \sin\phi_i\sin\varphi) \\ & + (ka)^2\frac{1}{45}\Bigg[(32-6\sin^2\theta_i)\cos\phi_i\cos\theta_i\cos\varphi \\ & + (32-10\sin^2\theta_i)\sin\phi_i\sin\varphi \\ & + \sin\vartheta\Big((6-8\sin^2\varphi)\cos\phi_i\cos\theta_i\sin\theta_i \\ & + 8\sin\phi_i\sin\theta_i\cos\varphi\sin\varphi\Big) \\ & - 6\sin^2\vartheta(\cos\phi_i\cos\theta_i\cos\varphi + \sin\phi_i\sin\varphi)\Bigg]\Bigg\}. \end{aligned} \tag{47b}$$

Here ϕ_i and θ_i are the angle of incidence as shown in Fig. 1. These results do not seem to agree with results obtained by Stevenson [24]. They lead, however, to the same expression for the scattering coefficient as obtained by Lur'e [19] and Kuritsyn [20].

VI. Scattering Coefficient for Incident Plane Wave

The scattering coefficient τ is defined as the ratio of the energy scattered from the disk to the energy incident on the disk. The scattered energy is found by integrating the Poynting vector of the scattered field over the sphere at infinity.

Using (47) one finds for an incident plane wave

$$\tau = \frac{\int_0^{2\pi}\int_0^{\pi} \mathrm{Re}\,(|H\vartheta|^2 + |H\phi|^2)\sin\vartheta d\vartheta d\varphi}{\pi a^2 H_0^2 \cos\theta_i}$$

$$= \frac{128}{27\pi^2\cos\theta_i}(ka)^4\left\{1+\sin^2\theta_i\left(\frac{5}{4}\sin^2\phi_i - 1\right)\right.$$

$$+\frac{(ka)^2}{25}[(22-5\sin^2\theta_i)\cos^2\theta_i\cos^2\phi_i$$

$$\left.+\tfrac{1}{4}(88-54\sin^2\theta_i-5\sin^4\theta_i)\sin^2\phi_i]\right\}. \quad (48)$$

For normal incidence Bouwkamp [23] obtained

$$\tau = \frac{128}{27\pi^2}(ka)^4\left[1+\frac{22}{25}(ka)^2\right.$$

$$\left.+\frac{7512}{18375}(ka)^4+0(ka)^6\right]. \quad (49)$$

A general expression for τ for arbitrary primary fields would involve the separation of the real and imaginary parts of the incident fields and their derivatives at the center of the disk. The resulting expression would be very cumbersome and of little practical value.

VII. Diffraction by a Circular Aperture

Using the results of the generalized Babinet's principle [51] it can be shown that the disk problem and the aperture problem are equivalent if the following substitutions are made

substitute $\mu_0 \boldsymbol{H}^i_{\text{Aperture}}$	for $\epsilon_0 \boldsymbol{E}^i_{\text{Disk}}$
substitute $-\epsilon_0 \boldsymbol{E}^i_{\text{Aperture}}$	for $\mu_0 \boldsymbol{H}^i_{\text{Disk}}$
substitute $\mu_0 \boldsymbol{M}_{\text{Aperture}}$	for $\boldsymbol{P}_{\text{Disk}}$
substitute $\boldsymbol{P}_{\text{Aperture}}$	for $\mu_0 \boldsymbol{M}_{\text{Disk}}$. (50)

It is customary to designate as primary field $\boldsymbol{E}^0$, $\boldsymbol{H}^0$ the field which would exist if the aperture is replaced by a solid screen. We obtain thus

$$\boldsymbol{E}_z^i = \tfrac{1}{2}\boldsymbol{E}_z^0$$
$$\boldsymbol{H}_{\tan} = \tfrac{1}{2}\boldsymbol{H}_{\tan}^0. \quad (51)$$

This leads to the following expressions for the induced electric and magnetic dipole moments:

$$P_z = \frac{4}{3}a^3\epsilon_0\left\{E_z^0 - \frac{1}{10}(ka)^2\left[3E_z^0 + \frac{1}{k^2}\frac{\partial^2 E_z^0}{\partial z^2}\right]\right.$$

$$\left.+j\frac{4}{9\pi}(ka)^3 E_z^0\right\}_0 \quad (52)$$

$$M_x = \frac{8}{3}a^3\left\{H_x^0 + \frac{(ka)^2}{30}\left[13H_x^0 - \frac{3}{k^2}\frac{\partial^2 H_x^0}{\partial z^2}\right.\right.$$

$$\left.\left.-\frac{2j}{\omega\mu_0}\frac{\partial E_z^0}{\partial y}\right] - j\frac{8}{9\pi}(ka)^3 H_x^0\right\}_0$$

$$M_y = \frac{8}{3}a^3\left\{H_y^0 + \frac{(ka)^2}{30}\left[13H_y^0 - \frac{3}{k^2}\frac{\partial^2 H_y^0}{\partial z^2}\right.\right.$$

$$\left.\left.+\frac{2j}{\omega\mu_0}\frac{\partial E_z^0}{\partial x}\right] - j\frac{8}{9\pi}(ka)^3 H_y^0\right\}_0. \quad (53)$$

The bracketed terms are evaluated at the center of the aperture.

The scattering coefficient t of an aperture is usually defined as the ratio between the energy incident on the aperture and the energy transmitted through the aperture. We obtain thus

$$t = \tfrac{1}{2}\tau \quad (54)$$

where τ is the scattering coefficient of a disk given by (48).

Appendix

Evaluation of Integrals

In the foregoing discussion we need to evaluate some integrals of the form

$$G(x,y) = \int_{\text{Disk}} \frac{f(x',y')}{\pi^2(a^2-x'^2-y'^2)^{1/2}((x-x')^2+(y-y')^2)^{1/2}}\,dS' \quad (55)$$

where $f(x', y')$ is a polynomial in x' and y'. Bouwkamp [50] has given a solution for this kind of integral in the form:

$$I(n,m,\mu;\rho,\varphi) = \int_0^a\int_0^{2\pi} \frac{P_{2n}^{2m}((a^2-\rho'^2)^{1/2})((x-x')^2+(y-y')^2)^{\mu/2}\cos(2m\varphi')}{(a^2-\rho'^2)^{1/2}((x-x')^2+(y-y')^2)^{1/2}}\rho' d\rho' d\varphi'$$

$$= \sum_v A_v(n,m,\mu)P_{2n}^{2m}((a^2-\rho^2)^{1/2})\cos(2m\varphi) \quad (56)$$

TABLE I

$$G(x, y) = \int_0^a \int_0^{2\pi} \frac{f(x', y')\rho' d\rho' d\varphi'}{\pi^2(a^2 - \rho'^2)^{1/2}[(x - x')^2 + (y - y')^2]^{1/2}}$$

$f(x, y)$	$G(x, y)$
1	1
x	$\frac{1}{2}x$
x^2	$\frac{1}{16}[4a^2 + 5x^2 - y^2]$
xy	$\frac{3}{8}xy$
x^3	$\frac{1}{32}[6xa^2 + 7x^3 - 3xy^2]$
x^2y	$\frac{1}{32}[2ya^2 - y^3 + 9x^2y]$
x^4	$\frac{1}{1024}[144a^4 + 144x^2a^2 - 48y^2a^2 + 169x^4 - 102x^2y^2 + 9y^4]$
x^3y	$\frac{1}{1024}[96xya^2 + 210x^3y - 60xy^3]$
x^2y^2	$\frac{1}{1024}[48a^4 + 16x^2a^2 + 16y^2a^2 - 17x^4 + 246x^2y^2 - 17y^4]$

TABLE II

$f(x, y)$	$\nabla^2 G(x, y)$
1	0
x	0
x^2	$\frac{1}{2}$
xy	0
x^3	$\frac{9}{8}x$
x^2y	$\frac{3}{8}y$
x^4	$\frac{3}{32}[2a^2 + 19x^2 - y^2]$
x^3y	$\frac{225}{256}xy$
x^2y^2	$\frac{1}{32}[2a^2 + 9x^2 + 9y^2]$

where

$\rho'^2 = x'^2 + y'^2 \leq a^2$
$\rho^2 = x^2 + y^2 \leq a^2$
$0 \leq \varphi \leq 2\pi$
P_{2n}^{2m} = associated Legendre functions
n, m π and μ = integers subject to the condition $\mu \geq 0$, $0 \leq m \leq n$.

The coefficients A_v are:

$$A_v(n, m, \mu) = \frac{\pi(-1)^{n+v}(2v + \frac{1}{2})\Gamma(\mu + 1)\Gamma\left(\frac{\mu + 1}{2}\right)\Gamma\left(\frac{\mu + 1}{2}\right)\Gamma(n + m + \frac{1}{2})\Gamma(v - m + \frac{1}{2})}{\Gamma(n - m + 1)\Gamma(v + m + 1)\Gamma(\frac{1}{2}\mu + n - v + \frac{1}{2})\Gamma(\frac{1}{2}\mu + n - v + 1)\Gamma(\frac{1}{2}\mu - n + v + 1)\Gamma(\frac{1}{2}\mu + n + v + 3/2)} \quad (57)$$

where $\Gamma(n+1) = n! =$ Gamma function

$$\Gamma(\tfrac{1}{2}) = (\pi)^{1/2}.$$

By choosing the integers n, m, and v appropriately and differentiating with respect to x and y the integral (55) can be found for any arbitrary polynomial $f(x, y)$. The calculations involved are, however, quite tedious.

Table I gives $G(x, y)$ for polynomials up to the fourth power in x and y.

Table II lists $\nabla^2 G(x, y)$.

Acknowledgment

The author is indebted to Dr. R. E. Collin for many valuable discussions on the subject of this paper. Thanks is due also to the Air Force Cambridge Research Center for sponsoring this work.

Bibliography

Scalar Diffraction Problem

[1] C. J. Bouwkamp, "Theoretische en numerieke behandeling van de buiging door een ronde opening," Ph.D. dissertation, University of Groningen, Germany; 1941.
[2] ——, "A Contribution to the Theory of Acoustic Radiation," *Philips Res. Rept.*, vol. 1, pp. 251–277; August, 1946.
[3] ——, "On the freely vibrating circular disk and the diffraction by circular disks and apertures," *Physica*, vol. 16, pp. 1–16; January, 1950.
[4] ——, "On the transmission coefficient of a circular aperture," *Phys. Rev.*, vol. 75, p. 1608; May, 1949.
[5] J. Bazer and A. Brown, "Diffraction of scalar waves by a circular aperture," IRE Trans. on Antennas and Propagation (Special Suppl., vol. AP-7, pp. S12–S20; December, 1959.
[6] H. Levine and J. Schwinger, "On the theory of diffraction by an aperture in an infinite plane screen," *Phys. Rev.*, vol. 74, pt. 1, pp. 958–974, October, 1948; vol. 75, pt. 2, pp. 1423–1432; November, 1949.
[7] H. Severin, "The Scalar Problem of Diffraction of Sound Waves by Thin Rigid Screens and its Application in Similar Vector Electromagnetic Problems in the Microwave Region," The McGill Symposium, AFCRC-TR-59-118 (II), Astia AD 211500, 256–263; 264–269; June, 1953.
[8] A. T. De Hoop, "On the scalar diffraction by a circular aperture in an infinite plane screen," *Appl. Sci. Res.*, vol. B4, pp. 151–160; September, 1954.
[9] W. Braunbeck, "Neue Näherungsmethode für die Beugung am ebenen Schirm," *Z. Phys.*, vol. 127, pp. 381–390; March, 1950.
[10] ——, "Zur Beugung an der Kreisscheibe," *Z. Physik.*, vol. 127, pp. 405–415; March, 1950

Vector Diffraction Problem

Rigorous Solution

[11] J. Meixner, "Strenge Theorie der Beugung elektromagnetischer Wellen an der vollkommen leitenden Kreisscheibe," *Z. Naturforsch.*, vol. 3a, pp. 506–518; August-November, 1948.

[12] J. Meixner and W. Andrejewski, "Strenge Theorie der Beugung ebener elektromagnetischer Wellen an der vollkommen leitenden Kreisscheibe und an der kreisförmigen Öffnung im vollkommen leitenden Schirm," *Ann. Phys.*, vol. 7, pp. 157–168; March, 1950.
[13] C. Flammer, "The vector wave function solution of the diffraction of electromagnetic waves by circular disks and apertures," I: "The oblate spheroidal vector wave function," *J. Appl. Phys.*, vol. 24, pp. 1218–1223, September, 1953; II: "The diffraction problem," pp. 1224–1231.
[14] K. Westpfahl, "Zur strengen Theorie der Beugung elektromagnetischer Wellen an ebenen Schirmen allgemeiner Gestalt," *Z. Physik.*, vol. 141, pp. 354–373; July, 1955.
[15] Y. Nomura and S. Katsura, "Diffraction of electromagnetic waves by circular plates and circular holes," *J. Phys. Soc. Japan*, vol. 10, pp. 285–304; April, 1955.
[16] N. I. Akhiezer and A. N. Akhiezer, "The problem of diffraction of electromagnetic waves at a circular aperture in a plane screen," *Doklady Akad. Nauk. S.S.S.R.*, vol. 109, pp. 53–56; January, 1956. (In Russian.)
[17] M. G. Belkina, "Diffraction of electromagnetic waves on a disk," Coll: "Diffraction of electromagnetic waves on certain figures of rotation," *Soviet Radio*, pp. 148–174; February, 1957. (In Russian.)
[18] N. N. Lebedev and I. P. Skal'skaya, "A new method for solving the problem of diffraction of electromagnetic waves by a thin conducting disk," *J. Tech. Phys. (U.S.S.R.)*, vol. 29, p. 700; 1959. (In Russian.) (English Translation in *Soviet Phys. JETP*, vol. 4, pp. 627–637; December, 1959.
[19] K. A. Lur'e, "Diffraction of a plane electromagnetic wave on an ideally conducting disk," *J. Tech. Phys. (U.S.S.R.)*, vol. 29, pp. 1421–1433; December, 1959. (In Russian.) (English Translation in *Soviet Phys. JETP*, vol. 4, pp. 1313–1325; June, 1960.)
[20] V. N. Kuritsyn, "Arbitrary incidence of a plane wave on a conducting disk," *J. Tech. Phys. (U.S.S.R.)*, vol. 30, pp. 790–798; 1960. (In Russian.) (English Translation in *Soviet Phys. JETP*, vol. 5, pp. 744–752; January, 1961.)

Approximate Solutions

Small Disks: ka < 1

[21] H. A. Bethe, "Theory of diffraction by small holes," *Phys. Rev.*, vol. 66, pp. 163–182; October, 1944.
[22] C. J. Bouwkamp, "On Bethe's theory of diffraction by small holes," *Philips Res. Repts.*, vol. 5, pp. 321–332; October, 1950.
[23] ——, "On the diffraction of electromagnetic waves by small circular disks and holes," *Philips Res. Repts.*, vol. 5, pp. 401–422; December, 1950.
[24] A. F. Stevenson, "Solution of Electromagnetic Scattering Problems as Power Series in the Ratio (Dimension of Scatterer/Wavelength)," The McGill Symposium, AFCRC-TR-59-118 (II), Astia AD 211500, pp. 235–238; June, 1953.
[25] G. A. Grinberg and Yu. V. Pimenov, "On the question of diffraction of electromagnetic waves from infinitesimally thin ideally conducting screens," *J. Tech. Phys. (U.S.S.R.)*, vol. 27, p. 2326; 1957. (In Russian.) (English Translation in *Soviet Phys. JETP*, vol. 2, pp. 2160–2175; October, 1957.)

Large Disks: ka > 1

[26] R. F. Millar, "An approximate theory of the diffraction of an electromagnetic wave by an aperture in a plane screen," *Proc. IEE*, Monograph No. 152 R, 9 pp.; October, 1955.
[27] ——, "The diffraction of an electromagnetic wave by a circular aperture," *Proc. IEE*, Monograph No. 196 R, 9 pp.; September, 1956.
[28] ——, "The diffraction of an electromagnetic wave by a large aperture," *Proc. IEE*, Monograph No. 213 R, 11 pp.; December, 1956.
[29] W. E. Frahn, "Beugung elektromagnetischer Wellen in Braunbekscher Näherung," *Z. Physik*, vol. 156, pp. 78–98, August, 1959; pt. 2, pp. 99–116, September, 1959.
[30] G. A. Grinberg and Yu. V. Pimenov, "Diffraction of electromagnetic waves by a circular aperture in an ideally conducting plane," *J. Tech. Phys. (U.S.S.R.)*, vol. 29, pp. 1206–1211; October, 1959. (In Russian.) (English Translation in *Soviet Phys. JETP*, vol. 4, pp. 1106–1111; April, 1960.
[31] S. R. Seshadri and T. T. Wu, "High-frequency diffraction of electromagnetic waves by a circular aperture in an infinite plane conducting screen," IRE Trans. on Antennas and Propagation, vol. AP-8, pp. 27–36; January, 1960.

Variational Methods

[32] J. W. Miles, "On electromagnetic diffraction through a plane screen," *Phys. Rev.*, vol. 75, pp. 695–696; February, 1949.
[33] ——, "On the diffraction of the electromagnetic wave through a plane screen," *J. Appl. Phys.*, vol. 20, pp. 760–771; August, 1949.
[34] H. Levine and J. Schwinger, "On the theory of electromagnetic wave diffraction by an aperture in an infinite plane conducting screen," *Comm. Pure & Appl. Math.*, vol. 3, pp. 355–391; December, 1950.

Experimental Work

[35] C. L. Andrews, "Diffraction pattern of a circular aperture at short distances," *Phys. Rev.*, vol. 71, pp. 777–786; June, 1947.
[36] ——, "The diffraction pattern of electromagnetic waves in the planes of apertures," *Phys. Rev.*, vol. 74, p. 1231; November, 1948.
[37] ——, "Diffraction pattern in a circular aperture measured in the microwave region," *J. Appl. Phys.*, vol. 21, pp. 761–767; August, 1950.
[38] R. D. Kodis, "Diffraction measurements at 1.25 centimeters," *J. Appl. Phys.*, vol. 23, pp. 249–255; February, 1952.
[39] H. L. Robinson, "Diffraction patterns in circular apertures less than one wavelength in diameter," *J. Appl. Phys.*, vol. 24, pp. 35–38; January, 1953.
[40] G. Bekefi, "Diffraction of Electromagnetic and Acoustic Waves by Apertures of Different Shapes," The McGill Symposium, AFCRC-TR-59-118 (II), Astia AD 211500, pp. 249–255; June, 1953.
[41] Chang Huang, "Diffraction by Apertures," The McGill Symposium, AFCRC-TR-59-118 (II), Astia AD 211500, pp. 270–279; June, 1953.
[42] A. Boivin, *et al.*, "Experimental study of the diffraction of microwaves by apertures with rotational symmetry," *Can. J. Phys.*, vol. 34, pp. 116–178; February, 1956. (In French.)

Other References

[43] F. Möglich, "Beugungserscheinungen an Körpern von ellipsoidischer Gestalt," *Ann. Phys.*, vol. 4, pp. 609–734; August, 1927.
[44] H. E. J. Neugebauer, "Diffraction of electromagnetic waves caused by apertures in absorbing plane screens," IRE Trans. on Antennas and Propagation, vol. AP-4, pp. 115–119; April, 1956.
[45] C. J. Bouwkamp, "Diffraction Theory, A Critique of Some Recent Developments," New York University, N. Y., Math. Res. Group, Res. Rept. No. Em-50; April, 1953.
[46] R. W. P. King and T. T. Wu, "The scattering and diffraction of waves," in "Harvard Monograph in Applied Science," Harvard University Press, Cambridge, Mass., no. 7, ch. 5, pp. 113–139; 1959.
[47] A. Sommerfeld, "Die Greensche Funktion der Schwingungsgleichung," *Jahresber. dtsch. Math.-Ver.*, vol. 21, pp. 309–353; March, 1913.
[48] J. Meixner, "Die Kantenbedingung in der Theorie der Beugung elektromagnetischer Wellen an vollkommen leitenden Schirmen," *Ann. Phys.*, vol. 6, pp. 1–9; January, 1949.
[49] C. J. Bouwkamp, "On the evaluation of certain integrals occuring in the theory of the freely vibrating circular disk and related problems," *Proc. Kon. Med. Akad. Wetensch.*, vol. 52, pp. 987–994; November, 1949.
[50] ——, "On integrals occurring in the theory of diffraction of electromagnetic waves by a circular disk," *Proc. Kon. Ned. Akad. Wetensch.*, vol. 53, pp. 654–661: May, 1950.
[51] R. E. Collin, "Field Theory of Guided Waves," McGraw-Hill Book Company, Inc., New York, N. Y.; 1960.
[52] W. H. Eggimann, "Higher Order Evaluation of the Diffraction of Electromagnetic Fields by Circular Disks," Case Inst. of Tech., Cleveland, Ohio, Sci. Rept. No. 21, AF 19(604)-3887; January, 1961.
[53] D. S. Jones, "A new method for calculating scattering with particular reference to the circular disc," *Comm. Pure and Appl. Math.*, vol. 9, pp. 713–746; November, 1956.

Low-Noise Properties of Microwave Backward Diodes*

SVERRE T. ENG†, MEMBER, IRE

Summary—This paper describes, for what is believed to be the first time, the low-noise properties of backward tunnel diodes in microwave applications. The physics of the diodes are reviewed together with some of the characteristics and equivalent circuit parameters. The diodes are then considered as mixer diodes with IF in the audio range and also the standard 30-Mc IF. Another promising application considered is the use of the backward diodes in low-level detection.

The results show that the noise figure at 13.5 kMc with a 1-kc IF is around 15 db better than any commercially available mixer diodes. Using 30-Mc IF, the noise figure of backward diode mixers is without special optimum design, comparable to the best mixer diodes on the market. Of great importance, especially in microminiaturization, is the fact that these diodes may be used with a very low local oscillator power (50 μw or less). The high nonlinearity of the I-V characteristic at the origin and the low 1/f noise properties of these diodes are also of benefit in crystal video receivers and other low-level detector applications.

I. Introduction

THERE are three basic types of noise which occur in semiconductor devices. Thermal noise is caused by random motion of electrons. When a drift velocity is superimposed by means of an electric field, we get shot noise. Then we have 1/f noise which is detected over and above thermal and shot noise, and which is distinguished by its spectral intensity.

The investigation reported in this paper is concerned with low 1/f noise (excess noise) and thermal noise in experimental samples of microwave backward diodes. The problem of noise reductions in mixer diodes is discussed in Section II, and the ideas which led to the fabrication of the backward diodes are also described.

Backward diodes made by diffusion or alloying techniques have not yet been accessible for microwave applications because of the large capacitances associated with the junctions. However, the pulse-bonding technique used in our experiments proves to be well adapted for microwave versions of these diodes. The physics of the diodes are also reviewed together with some of the characteristics and equivalent circuit parameters. These diodes have a nearly conventional diode forward I-V characteristic and the back characteristic of a tunnel diode.

A reduction of 1/f noise is of interest, for example, in Doppler radar systems where the IF is in the audio range. Since light weight, small size and simplicity are very important in this type of antenna-navigations radar, it does not seem desirable in some cases to increase the complexity of the system and go to higher IF (30 Mc) and thus possibly eliminate the 1/f noise problems. In some mixer diodes there is indication of a slight contribution of excess noise at 30 Mc. However, the main contribution to the noise at this frequency is shot and thermal noise. Because of the low spreading resistance in the semiconductor wafer, the thermal noise will be reduced somewhat compared with that of ordinary mixer diodes. Thus, an improvement of the sensitivity of systems using 30-Mc IF may also be expected. Present video crystal receivers and other low-level detector applications may also benefit in sensitivity from an 1/f noise reduction in diodes. All these application aspects of the backward diodes are described in the last part of this paper.

II. Reduction of Diode Noise

Although the origin of the 1/f noise in semiconductors is not precisely known, a variety of experiments have indicated that the surface plays a significant part in connection with this noise. It has been observed that 1/f noise increases with higher density of slow surface states, and, for a given density of slow states, the 1/f noise will be higher for an inversion layer than for an accumulation layer [1].

One approach for decreasing the 1/f noise is a stabilization of the surface by etching, ambient gas control, or thermally grown oxide layers. These techniques for stabilization and evaluation have been successfully applied to conventional mixer diodes in our laboratory.

However, the author is suggesting another approach to the problem of 1/f noise reduction in mixer diodes. For nearly degenerate material the relative change in surface potential is much less pronounced than in the nearly intrinsic case. Thus, if nonlinearities in the I-V characteristics of diodes made of much lower resistivity material than conventional mixer diodes were utilized for mixing purposes, the surface contribution to the 1/f noise should be reduced. Also, since the resistivity of the semiconductor wafer is reduced, the thermal noise contribution from the spreading resistance may be smaller.

The backward diodes used in our experiment have doping levels one and two orders of magnitude higher than conventional mixer diodes. Another point of interest is that the 1/f noise decreases when the rectified current decreases [2]. Since the conversion loss can be preserved at very low local oscillator power when using the backward diodes as mixers, improvements in the over-all noise figure of systems using IF frequencies in

* Received by PGMTT, March 28, 1961; revised manuscript received, June 9, 1961.

† Hughes Semiconductor Div., Hughes Aircraft Co., Newport Beach, Calif.

the audio range should be obtained. The reduction of the spreading resistance may also be of value in decreasing the noise figure of receiver systems using 30-Mc IF.

III. The Backward Diode [3], [4]

Since electrons can be considered to have wave properties, they have the ability to penetrate potential barriers which would be impossible for particles according to classic theory. This effect is known as quantum-mechanical tunneling. A backward diode is a special case of a diode exhibiting quantum-mechanical tunneling phenomena. The most familiar example is the tunnel diode, where the impurity concentrations on both sides of the p-n junction are such that an increase in tunneling current is followed by a decrease, producing a negative resistance. The expression for the tunneling current is

$$J \propto \exp\left[-\frac{Am^{*1/2}Eg^{3/2}}{F}\right], \tag{1}$$

where

A = numerical constant
m^* = reduced mass of holes and electrons
Eg = band gap of semiconductor
F = average electric field across the space-charge region.

F is inversely proportional to the width of the space-charge region of the p-n junction. A wide space-charge region means low tunneling current; a narrow one means high tunneling current. For a backward diode, it is necessary to reduce the tunneling current from the high levels of a tunnel diode to such levels that a negative resistance is either nonexistent or is so high that it can be disregarded. The expression for the width of the space-charge region of an abrupt junction is

$$W = \left[\frac{2K\,Ko(N_D + N_A)}{eN_DN_A}(V_D + V_A)\right]^{1/2}, \tag{2}$$

where

N_D = donor concentration
N_A = acceptor concentration
V_D = built-in "diffusion" voltage
V_A = applied bias voltage
$K\,Ko$ = dielectric constant of semiconductor
e = electronic charge.

If, in fabrication, one doping level is kept constant at some high level, the space-charge region can be widened by simply decreasing the impurity level of the semiconductor wafer. This causes a decrease of the tunneling current, and enables the device designer to adjust it to the proper level.

It is theoretically possible to make backward diode junctions by diffusion techniques. However, it is very impractical to control a diffusion of only several hundred atomic layers into the wafer. The technological difficulties of making electrical connections to such a shallow diffused layer are also severe. In addition, diffused junctions are undesirable because the tunneling probability decreases exponentially with the width of the space-charge region.

Tunnel diodes and also backward diodes are generally made by alloying techniques. An appropriate alloy pellet is placed in contact with a semiconductor wafer and heated to such a temperature that wetting and dissolutions of some of the semiconductor into the alloy occurs.

The alloy contains the impurity that will give opposite conductivity type to that in the wafer. On cooling the dissolved semiconductor will epitaxially grow on the base semiconductor. This regrowth will now be doped to the opposite conductivity type. There will be an abrupt transition between the doping levels on the two sides of the junction. The capacitance of this type of junction may be too large for practical use in the microwave region. However, by subsequent etching, the junction diameter can be reduced. The main advantage of this technique is that fairly good control of the current-voltage characteristic can be obtained. The disadvantages are that capacitances in the 0.2 $\mu\mu$f range are difficult to reproduce and that the structure is mechanically weak.

The pulse-bonding technique seemed to be the best way of fabricating our diodes. A gallium-plated gold wire was brought into contact with an n-type germanium semiconductor wafer with a resistivity in the range of 0.001 to 0.004 ohm/cm, and the junction was formed by passing a pulse through it. The pulsing equipment has certain parameters to vary, such as the heights and lengths of the pulse. The I-V characteristic of the diode can be displayed on a scope so that the junction can be pulsed to the desirable current-voltage characteristic. The capacitances obtained are well below one picofarad. The capacitance-voltage relationship is that of a step junction.

Unlike tunnel diodes, the package is not too critical. A standard computer glass package supplemented with gold-plated brass adaptors and with over-all dimensions confirming with ordinary mixer diode packages was found to be satisfactory at 13.5 kMc.

The equivalent circuit of the diode used in the experiment is shown in Fig. 1. C_t is the space-charge capacitance, which was around 0.2–0.5 $\mu\mu$f at zero bias. The series wafer resistance R_S was found to be approximately 10 ohms when estimated from the shape of the I-V curve in the forward direction. The lead inductance L_S is mμh when measured at 1 kMc, and C is the stray capacitance of the package. The variable resistance R_b varies between 300 ohms and 80 ohms for local oscillator powers in the range from 10 μw to 1 mw.

The nonlinear barrier resistance is believed to be independent of frequency from dc to well beyond the microwave region and is also considerably less temperature dependent than that of a conventional mixer diode.

The backward diode I-V characteristics are shown in Fig. 2. As the reverse bias is increased, the supply of electrons which are able to tunnel increases without limit. At zero bias no current flows through the junction. At forward bias only a very small current is able to flow by the tunneling process. The excess current is, in some cases, fairly constant until carrier injection takes place giving rise to the normal forward characteristic of a *p-n* junction.

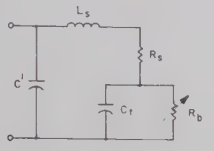


Fig. 1—The low-frequency diode equivalent circuit.

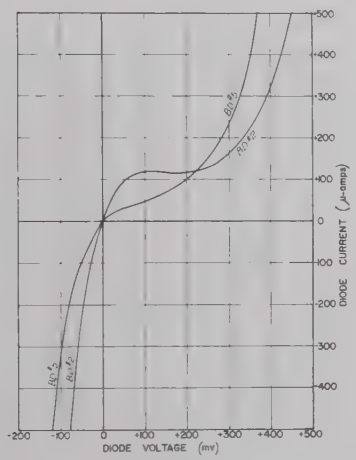


Fig. 2—The backward diodes' I-V characteristics.

IV. Application and Measurement

A. Mixer With IF in the Audio Range

Radars for navigation and moving target indications use the Doppler effect in one form or another. Since light weight, small size and simplicity are very important, it does not seem desirable in some cases to increase the complexity of the system by using 30-Mc IF to eliminate the 1/f noise problems. One simple arrangement is shown in Fig. 3. Power from the transmitter is mixed with the CW echo from a target. The beats between the two frequencies f and f' can be heard in phones or indicated on other types of indicators. The echo frequency is given by the well-known Doppler formula,

$$f' = \frac{c + v}{c - v} f. \tag{3}$$

The beat frequency is then

$$f_d = f' - f = \frac{2v}{c - v} \cdot f \simeq \frac{2v}{c} = \frac{2v}{\lambda}, \tag{4}$$

where v is the target velocity and λ is the wavelength of the transmitted signal. The beat frequency is usually in the kc range.

The operating noise figure of our Doppler system can be written as

$$F_{op} = (t_a - 1)\frac{B_t}{B_u} + L_m(t_m + F_{IF} - 1), \tag{5}$$

where

t_a = source temperature normalized with respect to room temperature
B_u = useful channel bandwidth
B_t = total channel bandwidth
L_m = mixer conversion loss
F_{IF} = IF amplifier noise figure.

The properties of the crystal diode itself are involved explicitly or implicitly in this equation. In our system the total bandwidth is approximately twice the useful bandwidth ($B_t/B_u \simeq 2$), since the signal is entering only one of the channels, while the antenna noise is appearing in the signal channel as well as the image channel.

Since 13.5 kMc is used in several types of navigation Doppler radar systems, the noise figures of the diodes were tested at this frequency in the experimental system shown in Fig. 4. In this way the noise figure can easily be measured with the so-called "signal generator

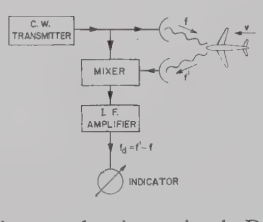


Fig. 3—Schematic diagram showing a simple Doppler radar system.

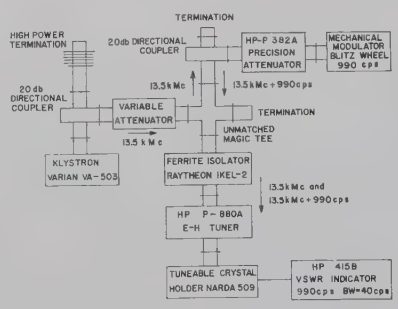


Fig. 4—The Doppler radar noise-figure measurement system.

method." The results obtained agree, within measurement errors, with calculated values of the noise figure obtained from separate measurement of the conversion loss and noise temperature. Microwave power at 13.5 kMc travels from the signal generator to the magic tee. One part of the power is delivered to the test diode and may be thought of as the local oscillator power of the mixer. Another part of the power travels through the precision attenuator to the "Blitz Wheel" mechanical modulator. The modulator reflects power at 13.5 kMc +990 cps, and part of this signal is available for delivery to the diode. This power may be considered the input signal, and it beats with the local oscillator signal in the test diode to produce the difference frequency of 990 cps, which is then amplified together with the audio noise generated by the diode. The calibration of the system is made by careful measurement of the individual insertion losses of the various microwave components and the response characteristic of the output meter. The signal amplitude is controlled by the precision attenuator in front of the modulator, and this input signal is adjusted so that the noise plus signal output power is double the output power obtained when the input signal is zero. Then the noise figure is obtained from the ratio of the input signal power used to the available thermal noise power from the source.

In order to properly evaluate the capabilities of the diodes, it was also found necessary to measure the conversion loss, noise temperature and the noise temperature spectrum of the experimental crystals [6].

The conversion loss L_m is defined as the ratio of the available RF input signal power to the measured IF output power at the mixer. In the "amplitude modulation method" used, the output of an oscillator at 13.5 kMc was amplitude modulated at 990 cps. The modulation envelope, after crystal detection, was developed across an output load. The conversion loss was then found from this voltage, since the percentage modulation and the power of the oscillator were known from careful calibration.

The mixer noise temperature t_m is defined as the ratio of available noise power output of the crystal to that of a resistor at room temperature. The "Y-factor method" [6] in which

$$t_m = F_{IF}(Y - 1) + 1 \tag{6}$$

was used in our case. Y is the ratio of system noise output with the test diode in the crystal holder to the noise output when a resistor equal to the IF dynamic resistance of the crystal was substituted for the crystal in the holder. The input unmodulated signal was at 13.5 kMc, and the crystal was followed by a 990-cps VSWR (bandwidth 40 cps, $F_{IF} \approx 6$ db) and a transistor post amplifier.

The noise temperature spectrum was measured under the same conditions as the noise temperature, except that the crystal was followed by a wide-band transistor amplifier and a spectrum analyzer.

In Fig. 5 the performance characteristics of a backward diode mixer with 990-cps IF are plotted. The noise figure of a commercially available low 1/f noise diode (1N1838) is also shown for comparison. First of all, it is indicated that the backward diode has a considerably lower noise figure which is caused by a reduction of 1/f noise (see Fig. 6). Another important consideration is that the diode is capable of operating with very low local oscillator powers without any dc bias. The reason is that the nonlinear region of the I-V characteristic is in the vicinity of the origin, while in ordinary mixer diodes the nonlinearity occurs around the contact voltage, which may be 0.3 to 0.6 volt in the forward direction. Thus, the backward diode may be used in a low-noise mixer using a tunnel diode or variable-capacitance diode harmonic generator as the local oscillator.

A more direct comparison of the two different diodes is made in Table I. Since the $R_S C_t$ product is higher in the BD No. 5 diode than in the 1N1838 diode, the conversion loss would be expected to be higher since the conversion loss is an increasing function of this product [6]. However, it should be pointed out that no real attempt has been made to optimize the fabrication technique.

Fig. 7 shows the noise temperature spectrums for the two diodes. It is interesting to note that the spectrums follow approximately the 1/f law. The turnover point from 1/f noise to shot and thermal noise is around 20 kc for the backward diode, and the noise temperature seems to be roughtly proportional to the square of the rectified current in the 1/f noise range. This current dependence of the noise temperature spectrum corresponds with results obtained for conventional point-contact mixer diodes [2].

It can be concluded that the real improvement in the noise figure obtained by the backward diode compared with that of a conventional mixer diode is caused by reduction in noise temperature. There is reason to believe that

1) 1/f noise is less pronounced in the backward diodes since they are made of lower resistivity material than ordinary mixer diodes.

2) 1/f noise is lower at the optimum noise figure because less local oscillator power (and thus rectified diode current) is needed for satisfactory operating performance.

B. Mixer with 30-Mc IF

In optimum receiver design it has been found that 30 Mc is a good choice for the intermediate frequency using present commercially available mixer diodes. This optimum IF is mainly determined by the noise figure of the IF amplifier and the mixer noise temperature. The

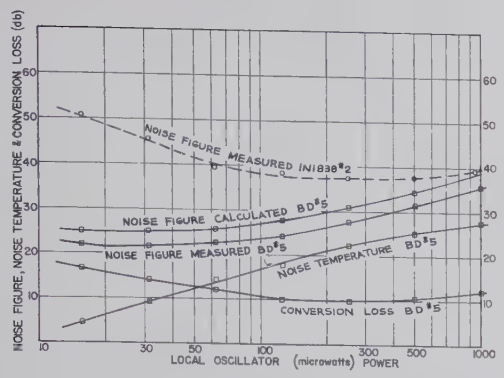


Fig. 5—Performance characteristics of a backward diode mixer with 990-cps IF. The noise figure of a 1N1838 diode (low 1/f noise) is also shown for comparison.

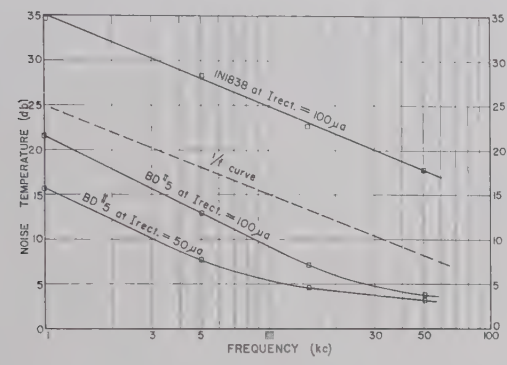


Fig. 6—The noise temperature spectrums for a backward diode and a 1N1838 low 1/f noise Doppler mixer diode.

TABLE I

Comparison of Mixer Characteristics
RF = 13.5 kMc, IF = 990 cps

Diode	R_S	C_t	30 μw Local Oscillator Power			250 μw Local Oscillator Power		
			F_{op}	L_m	t_m	F_{op}	L_m	t_m
	ohm	μμf	db	db	db	db	db	db
BD No. 5	10	0.3	21.5	14	9.8	27.5	10.0	21
1N1838	12	0.1	45.5	19.5	25	37.5	8	31.5

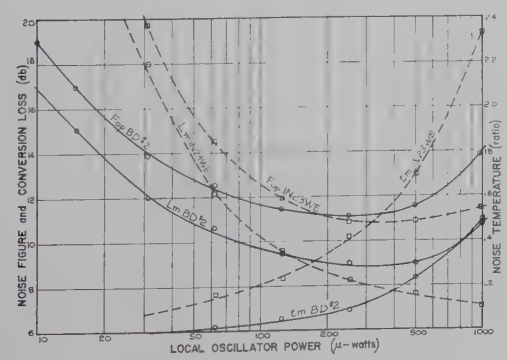


Fig. 7—Performance characteristics of a backward diode and a 1N23WE mixer diode with 30-Mc IF.

reduction of 1/f noise in mixer crystals may therefore influence receiver design.

At 30 Mc the main contribution to the noise temperature is shot and thermal noise. However, in some cases a very small excess noise contribution may be present. Nevertheless, since the backward diodes are made of lower resistivity material than conventional mixer diodes, a small improvement may be expected in the noise temperature since the thermal noise caused by the series bulk resistance will be reduced.

The experimental results are shown in Fig. 7. A comparison is made with the 1N23WE mixer diode. The noise figure was calculated according to (5) when $t_a = 1$, and the conversion loss was measured as outlined in Section IV-A. The noise temperature was measured with a 30-Mc IF amplifier ($F_{IF} = 1.7$ db) following the crystal.

In Table II the comparison is made at 300-μw local oscillator power, which seems to give the best noise figure for both diode types, and 30-μw, which may be of interest in lightweight systems since this small amount of power may be readily available from tunnel-diode oscillators or harmonic generators using variable-capacitance diodes.

TABLE II

Comparison of Mixer Characteristics
RF = 13.5 kMc, IF = 30 Mc

Diode	R_S	C_t	30-μw Local Oscillating Power			300-μw Local Oscillating Power		
			F_{op}	L_m	t_m	F_{op}	L_m	t_m
	ohm	μμf	db	db	ratio	db	db	ratio
BD No. 2	8	0.6	14.1	12.3	1	11.1	8.8	1.12
1N23WE	35	0.2	20	18	1.08	10.8	8	1.46

The backward diode is slightly higher in noise figure than the 1N23WE under conventional local oscillator power requirements. The difference is mainly caused by the conversion loss. This was expected, since the $R_S C_t$ product is highest for the backward diode. However, the 1N23WE seems to have a lower nonlinearity coefficient for the barrier resistance than the backward diode (this can be verified by inspection of the I-V curves for the respective diodes). Also, since the backward-diode noise temperature is, indeed, lower, there may be reason to believe that, by optimizing the fabrication technique so that a lower value of capacitance can be obtained, the backward diode may be superior to the 1N23WE in this application with 300-μw local oscillator power.

At lower local oscillator powers the backward diode has much better noise figure than the 1N23WE diode. The main reason for this is that the 1N23WE diode does not have the same ability to rectify small powers as the

backward diodes. Thus, the conversion loss will be substantially higher in the 1N23WE diode at low local oscillator powers.

C. Low-Level Detection

Another major application of the backward diode is in video detection. In a crystal video receiver, for example, the incoming modulated RF signal is detected immediately at the input from the antenna, and the resulting video signal is amplified by a high-gain video amplifier. The advantages of this type of receiver are simplicity, small size, low cost and broad RF bandwidth. The price paid for these attractive features is a large loss of sensitivity compared with that obtainable with a superheterodyne receiver. However, there are applications where the lower sensitivity is acceptable and where the small size or broad band is required (beacon radar receivers, square-law detectors in power measurement instruments, etc.).

The SNR of a detector crystal is proportional to a figure of merit [6] M, which combines the rectification properties and the noise generation of the crystal and the amplifier. A higher figure of merit indicates a greater sensitivity and thus a lower minimum detectable signal.

$$M = \frac{\beta R_b}{\sqrt{R_b + R_A}}, \tag{7}$$

where

β = short-circuit current sensitivity of the diode
R_b = video impedance of the diode
R_A = equivalent noise resistance of the video amplifier.

The current sensitivity β was measured at 13.5 and 6 kMc by observing the short-circuit rectified current i for a given input power P. Then $\beta = i/P$. The video resistance R_b is the dynamic dc resistance of the crystal at the operating point, and it was measured with a bridge at 100 kc.

The short-circuit current is plotted in Fig. 8 as a function of the absorbed input power at 13.5 and 6 kMc for a backward diode and a detector diode MA408B. The backward diode has a higher current sensitivity and a square-law response of a larger range than the conventional detector diode. A more direct comparison is made in Table III.

At 13.5 kMc the backward diodes seem more advantageous according to the figure of merit. The reason for the higher current sensitivity is partly due to the lower video impedance R_b compared with that of the MA408B diode. However, the product $R_S C_t$ is lower and the I-V nonlinearity is also higher for the backward diode. Both factors will result in higher current sensitivity. At 6 kMc the MA408B shows a higher figure of merit which may be contributed by the higher video resistance. There are good reasons for believing that the video resistance of the backward diode may be increased by suitable optimization of the shape of the I-V curve.

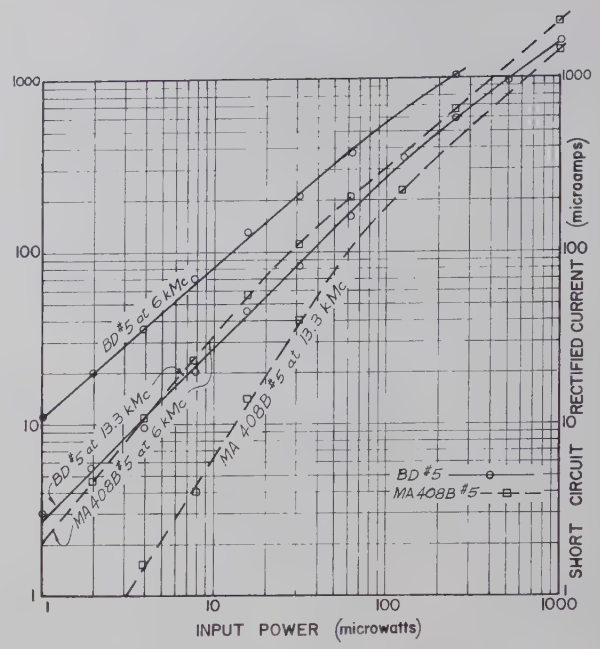


Fig. 8—The short-circuit current as a function of the absorbed input power for a backward diode and a detector diode MA408B.

TABLE III
COMPARISON OF VIDEO DETECTOR CHARACTERISTICS

Diode	R_S	C_t	R_b	13.5 kMc		6 kMc	
				β	M	β	M
	ohm	$\mu\mu$f	ohm	at 1 μw power	calculated	at 1 μw power	calculated
MA408B No. 5	42	0.23	35k	0.2	37	2	370
MA408B No. 6	32	0.32	20k	0.2	28	2	280
BD NO. 5	10	0.3	730	2.7	45	11	180
BD NO. 7	8	0.5	880	2.4	47	7.5	145

Other possible advantages of the backward detector diodes compared with conventional detector diodes may be the expected smaller variations in the performance characteristics with variations in temperature and nuclear radiation.

V. Conclusion

Experimental microwave backward diodes have been made, and the low-noise properties of these diodes investigated and compared with the best commercially available crystals. The performance characteristics show that backward diode mixers are of considerable advantage in receiver systems where the IF is in the audio range. The improvement is mainly caused by a

reduction in 1/f noise. It should be noted that a backward diode mixer can be operated with an order of magnitude lower local oscillator power than mixer diodes used today. Very satisfactory performance was also obtained using 30-Mc IF and in low-level detection, in spite of the fact that no real attempt was made to optimize the diode fabrication technique. In fact, there are good theoretical reasons for believing that other wafer materials may well prove to be superior to germanium for the applications considered in the paper. Since the backward diodes are virtually independent of the lifetime of minority carriers or of the surface treatment, they can tolerate larger doses of nuclear radiation than conventional mixer diodes. The tunneling portion of the I-V curve is also expected to be substantially independent of temperature.

ACKNOWLEDGMENT

The author wishes to thank D. English and R. Knox for help in performing the measurement, R. August and Dr. J. Morgan for help in fabricating the diodes, and Dr. E. L. Steele for permission to carry out this investigation.

REFERENCES

[1] A. U. MacRae and H. Levinstein, "Surface-dependent 1/f noise in germanium," *Phys. Rev.*, vol. 119, pp. 62–69; July, 1960.
[2] A. van der Ziel, "Noise," Prentice-Hall, Inc. New York, N Y., p. 220; 1954.
[3] I. A. Lesk, *et al.*, "Germanium and silicon tunnel diodes—design, operation, and application," 1959 IRE WESCON CONVENTION RECORD, pt. 3, pp. 9–31.
[4] R. N. Hall, "Tunnel diodes," IRE TRANS. ON ELECTRON DEVICES, vol. ED-7, pp. 1–9; January, 1960.
[5] D. O. North and H. T. Friis, "Discussion on noise figure of radio receivers," PROC. IRE, vol. 33, pp. 125–127; February, 1945.
[6] H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," McGraw-Hill Book Co., Inc., New York, N. Y.; 1948.

Design Theory of Up-Converters for Use as Electronically-Tunable Filters*

GEORGE L. MATTHAEI†, MEMBER, IRE

Summary—**The up-converters discussed use a single diode, a wide-band impedance matching filter at their signal input, a moderately wide-band impedance matching filter at their pump input, and a narrow-band filter at their sideband output. With a narrow-band filter at the sideband output, the frequency which will be accepted by the amplifier can be controlled by varying the pump frequency. Analysis of the impedance matching problem involved shows that tuning ranges of the order of a half-octave to an octave are possible. Theory is presented for both the lower-sideband and upper-sideband types of tunable up-converters and for the design of the required impedance-matching networks. It is shown that, because of the pump input bandwidth required, it will generally be necessary to accept some mismatch at the pump input. But, by use of a properly designed impedance-matching filter, the reflection loss can be kept nearly constant across the pump band, and the incident pump power required is not unreasonable. It is seen that properly designed devices of this type using voltage-tunable pump oscillators should have wide tuning range, fast tuning capability, a useful amount of gain, no image response, and a low noise figure.**

* Received by the PGMTT, April 17, 1961; revised manuscript received, June 19, 1961. The research reported here was sponsored by the U. S. Army Signal Res. and Dev. Lab., Ft. Monmouth, N. J., under Contract DA 36-039 SC-74862.
† Stanford Res. Inst., Menlo Park, Calif.

I. INTRODUCTION

A. Description of the Proposed Devices

PREVIOUS work dealt with the application of filter theory to the design of wide-band parametric amplifiers and up-converters.[1] The present discussion applies a similar theoretical approach to a different but closely related problem. The objective will be to obtain an electronically controlled *wide tuning range* using up-converters having a wide-band input-impedance-matching filter, a narrow-band output-impedance-matching filter, and a voltage-tunable pump oscillator such as a backward-wave oscillator.

Defining f as the input frequency, f' as the sideband output frequency, and f_p as the pump frequency, for a lower-sideband up-converter the output is at the lower-

[1] G. L. Matthaei, "A study of the optimum design of wide-band parametric amplifiers and up-converters," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-9, pp. 23–28; January, 1961.

sideband frequency

$$f' = f^p - f. \tag{1}$$

For an upper-sideband up-converter, the output is at the upper-sideband frequency

$$f' = f^p + f. \tag{2}$$

For either type of up-converter, tuning action can be achieved if a narrow-band filter is used at the output so that only frequencies equal or very nearly equal to a specific frequency f_0' can be passed. If the pump frequency f^p is varied, then the input frequency which will be accepted by the amplifier will be given by

$$f = f^p - f_0' \tag{3}$$

for the case of lower-sideband up-converters, and

$$f = f_0' - f^p \tag{4}$$

for the case of upper-sideband up-converters. In both cases, the amplifiers will yield gain. The lower-sideband type introduces some negative-resistance amplification in addition to the up-converter amplification,[2,3] however, and will therefore generally give more gain. For gain to be achieved, the variable-capacitance diode must see proper terminations at both frequencies, f and f_0'. Since f_0' is a fixed frequency, it should be relatively easy to maintain proper termination at that frequency. The input frequency f varies, however, and the tuning range of the amplifier will be determined largely by the range of f for which proper terminating conditions can be maintained.

B. Factors Permitting Large Tuning Range

For convenience, let us first consider the case of an upper-sideband up-converter with a fixed pump frequency. We analyze a wide-band, upper-sideband up-converter using a fixed pump frequency f^p, with an input-impedance-matching filter whose pass band is centered at f_0 and an output-impedance-matching filter whose pass band is centered at f_0'. In order to obtain maximum operating bandwidth, it is necessary that the condition

$$\frac{w}{w'} = \frac{f_0'}{f_0} \tag{5}$$

be satisfied, where w is the fractional bandwidth $\Delta f/f_0$ of the input filter and w' is the fractional bandwidth $\Delta f'/f_0'$ of the output filter.[1] Eq. (5) says, in effect, that the input and output bands Δf and $\Delta f'$, respectively, must be equal. From the theory of wide-band up-converters,[1] it is seen that if, for example, Δf were made larger than $\Delta f'$, the output filter bandwidth $\Delta f'$ would have to be shrunk to be *smaller* than it could be if Δf and $\Delta f'$ were made to be equal. Since the smaller bandwidth will be the one that limits the over-all transmission, the operating bandwidth necessarily becomes smaller if Δf and $\Delta f'$ are made unequal. Thus it is seen that for the usual case where the pump frequency is fixed, wide-band impedance matching must be accomplished both at the signal input and also at the upper-sideband output band, and best results are obtained if the bandwidth for impedance match is equal for both of these bands.

For the tunable up-converters considered herein, a quite different situation exists. In this case, narrow rather than wide bandwidth is desired. A wide-band impedance match is desired at the input channel, however, so that the input acceptance band can be moved about by varying the pump frequency. Under these conditions, we may expand the input bandwidth Δf and be quite happy with the required shrinkage of the sideband output bandwidth $\Delta f'$. With such an amplifier, the bandwidth $\Delta f'$ of the upper-sideband filter will be the operating bandwidth of the amplifier, while Δf will become the tuning range. If there were no practical considerations, such as the effects of losses and the need for practical impedance levels, the tuning range Δf could be made to approach infinity while the output bandwidth $\Delta f'$ would approach zero; practical considerations, however, do limit the obtainable tuning range. Nevertheless, it is clear that the *tuning range* obtainable with such a device is considerably greater than the *bandwidth* obtainable in an up-converter using a fixed pump frequency. As is seen in Matthaei,[1] bandwidths of the order of 16 per cent appear to be feasible for upper-sideband up-converters having fixed pump frequencies. The tuning range which is feasible for up-converters with varying pump frequencies is thus likely to be several times this value.

An additional problem, not yet mentioned, arises in this case of an up-converter with varying pump frequency. Additional impedance matching for the pump input channel must be provided in order to pump the diode satisfactorily at frequencies throughout the required pump frequency range. It can be shown that it will generally be impossible to obtain a good match between the pump signal oscillator and the small resistance in the diode while tuning across the required pump frequency band. However, it is possible to design impedance-matching filters which will give a nearly uniform, minimum reflection level throughout the required band. Then it is merely necessary to have a pump oscillator with sufficient available power to compensate for the reflection loss.

The discussion above has been phrased in terms of upper-sideband up-converters. The same principles,

[2] J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements—part 1. General energy relations," PROC. IRE, vol. 44, pp. 904–913; July, 1956.

[3] H. E. Rowe, "Some general properties of nonlinear elements. II. Small signal theory," PROC. IRE, vol. 46, pp. 850–860; May, 1958.

however, apply to lower-sideband up-converters. The main difference between the two devices is that the lower-sideband type tends to reflect a negative resistance through the coupling action of the varying capacitance. This negative resistance has the advantage that it contributes to the gain of the amplifier, but it also has the disadvantage that it could lead to oscillation if the amplifier were improperly terminated. Calculations indicate that good results should be obtainable by keeping the negative resistance small, so that stability will not be a problem.

II. Definition of Diode Parameters

The specification of the diode parameters will be the same as that used previously,[1,3] but will be summarized here for easy reference. The diode capacitance parameters C_0 and C_1 are defined by

$$C(t) = C_0 + 2C_1 \cos(2\pi f^p t + \phi_1) + \cdots, \tag{6}$$

where $C(t)$ is the time variation of the pumped diode. If the diode is to be resonated in shunt, it is convenient to use the short-circuit admittances[1,3]

$$\begin{aligned} B_{11} &= 2\pi f C_0, & B_{12} &= 2\pi f C_1 \\ B_{21} &= 2\pi f' C_1, & B_{22} &= 2\pi f' C_0. \end{aligned} \tag{7}$$

In the discussion to follow, subscript zeros, such as

$$(B_{11})_0, \quad (B_{12})_0, \quad (B_{21})_0, \quad (B_{22})_0,$$

will be used to indicate parameters evaluated at the signal input midband frequency f_0, or at the sideband output midband frequency f_0', whichever is appropriate for the given parameter.

The equivalent shunt diode loss conductances are defined as

$$G_d = \frac{2\pi f_0 C_0}{Q_d} = \frac{(B_{11})_0}{Q_d} \tag{8}$$

seen by input signal components and

$$G_d' = \frac{2\pi f_0' C_0}{Q_d'} = \frac{(B_{22})_0}{Q_d'} = G_d \left(\frac{f_0'}{f_0}\right)^2 \tag{9}$$

seen by the sideband components. Q_d is the operating Q of the diode at frequency f_0, and Q_d' is the operating Q at f_0'. Since the diode loss resistance is effectively in series with the capacitance,

$$Q_d' = Q_d \frac{f_0}{f_0'}. \tag{10}$$

Analogously, if the diode is to be resonated in series, it is convenient to use the open-circuit impedances[1]

$$\begin{aligned} X_{11} &= \frac{1}{2\pi f C_0^s}, & X_{12} &= \frac{1}{2\pi f' C_1^s} \\ X_{21} &= \frac{1}{2\pi f C_1^s}, & X_{22} &= \frac{1}{2\pi f' C_0^s} \end{aligned} \tag{11}$$

where

$$C_0^s = C_0(1 - a^2), \qquad C_1^s = \frac{C_0(1 - a^2)}{a} = \frac{C_0^s}{a} \tag{12}$$

and

$$a = \frac{C_1}{C_0}. \tag{13}$$

In this case, the effects of losses are represented by series resistors

$$R_s = R_s' = \frac{1}{Q_d 2\pi f_0 C_0^s} = \frac{(X_{11})_0}{Q_d} = \frac{(X_{22})_0}{Q_d'}, \tag{14}$$

which are the same at both the signal and sideband frequencies.

III. Definition of Filter Parameters

In order to reduce the number of degrees of freedom involved in the amplifier design, the design of the various band-pass filters required will be based on the use of lumped-element, low-pass prototypes. When these low-pass prototypes have been specified, the only parameters which remain to be specified in the corresponding band-pass filters are their center frequencies, their impedance levels, and their fractional bandwidths. Their center frequencies, of course, will be determined by the desired operating bands. The impedance level of the input and pump channel filters will be determined from considerations of their desired bandwidths and the impedance-matching properties which are required.

Fig. 1 shows a typical low-pass prototype and a typical Tchebycheff response for such a filter. Tables of normalized element values for Tchebycheff and maximally flat low-pass filters of this kind are presently available[4,5] Fig. 2 shows the corresponding band-pass filter obtained by the standard low-pass to band-pass transformation given in Fig. 2(a).

In an actual microwave device it will usually be necessary to use resonators which are a combination of semi-lumped and transmission-line elements rather than lumped-element resonators such as are shown in Fig. 2. For this reason, it is convenient to use the equivalent representation shown in Fig. 3. In this case, the actual nature of the resonators is left unspecified, but it is assumed that in the vicinity of f_0 they exhibit resonance characteristics similar to those of the resonators in Fig. 2 [as is implied by Fig. 3(a) and (d)]. Then the properties of the individual resonators are specified in terms of the resonant frequency f_0 and a *slope parame-*

[4] L. Weinberg, "Additional tables for design of optimum ladder networks, Pts. 1 and 2," *J. Franklin Inst.*, vol. 264, pp. 7–23, 127–138; July, August, 1957.

[5] W. J. Getsinger, *et al.*, "Research on Design Criteria for Microwave Filters," Stanford Res. Inst., Menlo Park, Calif., Tech. Rept. No. 4, SRI Project 2326, Contract DA 36-039 SC-74862; 1958.

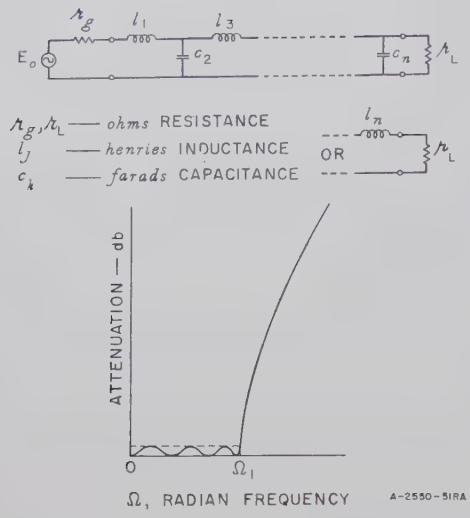


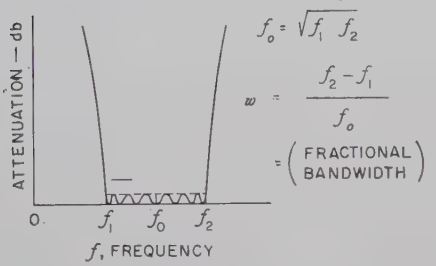
Fig. 1—Low-pass prototype filter and a typical Tchebycheff response.

$$\Omega = \frac{\Omega_1}{w}\left(\frac{f}{f_o} - \frac{f_o}{f}\right) \quad \text{(a)}$$

WHERE

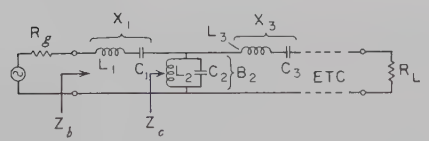
$$f_o = \sqrt{f_1 f_2} \quad \text{(b)}$$

$$w = \frac{f_2 - f_1}{f_o} = \left(\text{FRACTIONAL BANDWIDTH}\right) \quad \text{(c)}$$



IN EQUATIONS FOLLOWING,

$$\omega_o = 2\pi f_o \quad \text{(d)}$$



R_L IS CHOSEN TO GIVE DESIRED IMPEDANCE LEVEL, AND $R_g = R_L\,(r_g/r_L)$

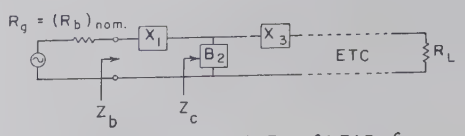
FOR SERIES BRANCHES

$$L_j = \left(\frac{R_L}{w\,\omega_o}\right)\left(\frac{\Omega_1 l_j}{r_L}\right), \quad C_j = \left(\frac{w}{\omega_o R_L}\right)\left(\frac{r_L}{\Omega_1 l_j}\right) \quad \text{(e)}$$

FOR SHUNT BRANCHES

$$C_k = \left(\frac{1}{w\,\omega_o R_L}\right)(\Omega_1 c_k r_L), \quad L_k = \left(\frac{w R_L}{\omega_o}\right)\left(\frac{1}{\Omega_1 c_k r_L}\right) \quad \text{(f)}$$

Fig. 2—Summary of relations for design of lumped-element band-pass filters from low-pass prototypes.



FOR SERIES BRANCHES, f NEAR f_o

$$X_j \approx \left(\frac{f}{f_o} - \frac{f_o}{f}\right) x_j \quad \text{(a)}$$

WHERE

$$x_j = (\text{SLOPE PARAMETER}) = \frac{\omega_o}{2}\frac{dX_j}{d\omega}\bigg|_{\omega=\omega_o} = \frac{R_{bo}}{w}\left(\frac{\Omega_1 l_j}{r_L}\right) \quad \text{(b)}$$

AND

$$R_{bo} = Z_b\big|_{f=f_o} \quad \text{(c)}$$

FOR SHUNT BRANCHES, f NEAR f_o

$$B_k \approx \left(\frac{f}{f_o} - \frac{f_o}{f}\right) b_j \quad \text{(d)}$$

WHERE

$$b_k = (\text{SLOPE PARAMETER}) = \frac{\omega_o}{2}\frac{dB_k}{d\omega}\bigg|_{\omega=\omega_o} = \frac{1}{w R_{bo}}(r_L \Omega_1 c_k) \quad \text{(e)}$$

Fig. 3—General description of band-pass filters in terms of resonator slope parameters (Ω_1, l_j, c_k and r_L are defined in Fig. 1).

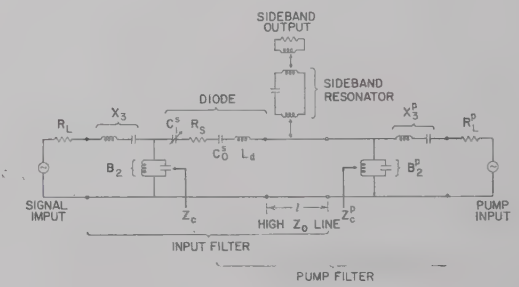


Fig. 4—A possible circuit for an up-converter for electronically-tunable filter applications.

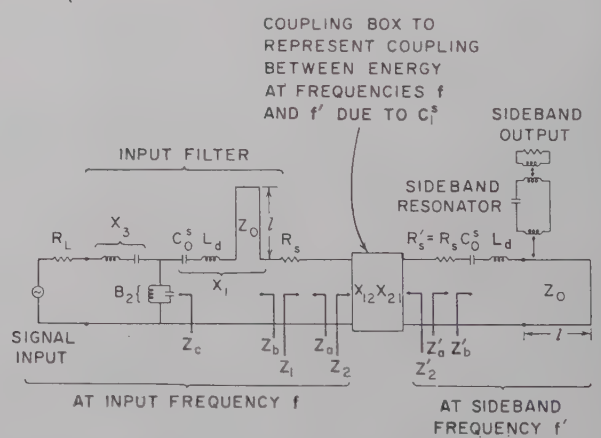


Fig. 5—A circuit which is approximately equivalent to that in Fig. 4 for energy components at the input frequency f or the sideband frequency f'.

ter. For a series, lumped-element resonator, the *reactance slope parameter* is simply

$$X_j = \omega_0 L_j = \frac{1}{\omega_0 C_j} = \frac{\omega_0}{2}\frac{dX_j}{d\omega}\bigg|_{\omega_0}, \tag{15}$$

where $\omega_0 = 2\pi f_0$ and X_j is the reactance of the resonator circuit. The derivative form in (15) and in Fig. 3(b) gives the corresponding reactance slope parameter for *any* series-resonant circuit, whether of lumped-element form or not. As indicated in Fig. 3(e), for circuits exhibiting a shunt resonance an analogous *susceptance slope parameter* b_k applies. Fig. 3(b) and (e) relates the band-pass filter resonator slope parameters to the elements of the low-pass prototype.

The parameter R_{b0} defined in Fig. 3(c) equals R_L for the circuit as shown. However, the definition given for R_{b0} is introduced because some microwave filters give impedance-level transformations so that R_{b0} may not necessarily equal R_L.

IV. An Up-Converter Model for Purposes of Analysis and Discussion

The up-converter model discussed in this section incorporates some features which have been used at SRI quite successfully on a degenerate parametric amplifier, and it is believed that they could also be used to advantage in electronically tunable up-converters in certain frequency ranges. However, the design equations and charts in the following sections should be applicable for any of a variety of coaxial, stripline, or waveguide types of up-converter structures. In this model, the diode will be resonated in series, although the design data to follow can be applied also for designs where the diode is resonated in shunt.

Fig. 4 shows the model up-converter circuit. This circuit uses a three-resonator wide-band input filter, a three-resonator moderately wide-band pump circuit filter, and a narrow-band, single-resonator output filter. The diode plus the length of high-impedance transmission line introduces multiple resonances, one of which provides the resonance for the No. 1 resonator of the input filter, and another of which provides the resonance for the No. 1 resonator of the pump filter. Note that the diode representation includes C_0^s, C_1^s, and R_s in series along with L_d, the diode internal inductance. Since the number of resonators for the signal-input and pump filters may vary under different circumstances, it is convenient for both the input and pump filters to define the resonator formed by the diode circuit as the No. 1 resonator, even though the generator is actually at the other end. Under these conditions, R_L becomes the generator internal impedance instead of R_g, but the power transmission properties are unaffected since the filter circuit obeys the reciprocity theorem.

The performance of the circuit in Fig. 4 can best be understood by considering its operation at the various frequencies of interest. Fig. 5 shows the equivalent circuit of the input filter circuit at frequency f and the output circuit at the sideband frequency f', with a box labeled $X_{12}X_{21}$ to represent the coupling effect of C_1^s between energy components at frequencies f and f'. The operation of this box is different for lower-sideband up-converters than for upper-sideband up-converters,[1] and its operation will be summarized in Sections V and VII. Since the pump filter center frequency f_0^p is much higher than the center frequency f_0 of the input filter, the susceptance B_2^p of the No. 2 resonator of the pump filter in Fig. 4 will be very large, so that the high-impedance line is effectively short-circuited. At frequency f_0, the high-impedance line is less than $\lambda/8$ long and acts much like a lumped inductance which brings the diode to series resonance at that frequency. Since C_0^s, L_d, R_s, the impedance of the short-circuited line, and the impedance reflected through C_1^s (Z_2 in Fig. 5) are effectively in series, the diode and high Z_0 line together may be represented as shown in Fig. 5 by R_s, Z_2, and the series resonator X_1.

At the pump-channel band-center frequency f_0^p, the impedance Z_c in Fig. 4 will be reactive and nearly zero, and the diode plus the high Z_0 line will again exhibit resonance. The impedance seen looking into the high Z_0 line from the right in Fig. 4 is given by

$$Z = \frac{(R_s + jX_d) + jZ_0 \tan\theta}{1 + \dfrac{j(R_s + jX_d)\tan\theta}{Z_0}}, \tag{16a}$$

where $R_s + jX_d$ is the total impedance developed across the diode along with the input filter reactance Z_c in Fig. 4 at the pump frequency f^p. Choosing Z_0 to be large as compared to R_s (a condition which is easy to fulfill since R_s might typically be around 4 ohms), to a good approximation (16a) becomes

$$Z = R_0^p + jX_1^p \approx \frac{R_s}{\left(1 - \dfrac{X_d}{Z_0}\tan\theta\right)} + \frac{j(Z_0\tan\theta + X_d)}{\left(1 - \dfrac{X_d}{Z_0}\tan\theta\right)}. \tag{16b}$$

In most cases, the pump band center f_0^p will be above the self-resonant frequency of the diode; therefore, at f_0^p the reactance X_d will usually be inductive, and θ will be somewhat less than π. Under these conditions, the quantity in the denominators of (16b) will be somewhat greater than one (since $\tan\theta$ is negative), and will be slowly varying since X_d/Z_0 will be increasing with frequency, while $\tan\theta$ will be decreasing in magnitude as the frequency increases.

Since θ will be roughtly of the order of π, in some cases the denominators of the terms in (16b) will be very nearly one. Assuming for the moment that this is the case and that Z_e for the input filter in Fig. 4 is negligible at $f_0{}^p$, then (16b) is equivalent to the impedance of R_s, $C_0{}^s$, L_d, and the short-circuited, high Z_0 line all connected in series as shown in Fig. 6(a). If the denominator in (16b) cannot be neglected, the magnitude of $R_e{}^p + jX_1{}^p$ will be affected as indicated by the equation, but the general nature of the performance will be qualitatively the same.

It is thus seen that for frequencies in the vicinity of f_0 and $f_0{}^p$, $C_0{}^s$, L_d and the high Z_0 line operate at least qualitatively in a manner similar to the series-resonant circuit shown in Fig. 6(a), which has the reactance characteristics shown in Fig. 6(b). The reactance slope in the vicinity of f_0 will be important in the design of the signal-input filter, while the reactance slope in the vicinity of $f_0{}^p$ will be important in the design of the pump input filter. At the sideband frequency f_0', the diode and high Z_0 line in Fig. 4 will see reactive impedances Z_e and $Z_e{}^p$ of small magnitude, but the diode and line will not be resonant. However, the presence of the lightly-coupled sideband resonator in Fig. 4 will introduce a narrow resonance at either the lower- or upper-sideband frequencies indicated in Fig. 6(b). It is desirable that this sideband-output resonator be coupled as lightly as possible, consistent with low transmission loss, so that it will cause a minimum increase in the reactance slope in the vicinity of $f_0{}^p$. Since the up-converter would probably be followed by a fixed-tuned, superheterodyne receiver, an additional high-Q filter following the sideband-output filter might be desirable to ensure high attenuation at the image frequency of the superheterodyne receiver.

Fig. 7 shows portions of a possible strip transmission line amplifier of the type under consideration. The diode used is in a computer diode package having a glass envelope and wire leads. (The Hughes 1N896 diodes are examples of varactor diodes packaged in this manner.) The wire diode leads then serve as the high Z_0 line for resonating the diode. Resonators Nos. 2 and 3 for the signal-input filter are realized as two semi-lumped shunt resonators separated by a quarter-wavelength (at frequency f_0) of line. The quarter-wavelength line eliminates the need to construct resonator No. 3 as a series resonator. The signal-input filter would be of wide bandwidth and it could be designed using the theory in Sections V or VII along with techniques for design of wide-band filters which have been previously treated.[6]

The pump frequency must be able to shift an amount equal to the bandwidth of the signal-input filter, but since the pump frequency band is centered considerably higher than that of the signal-input band, the pump

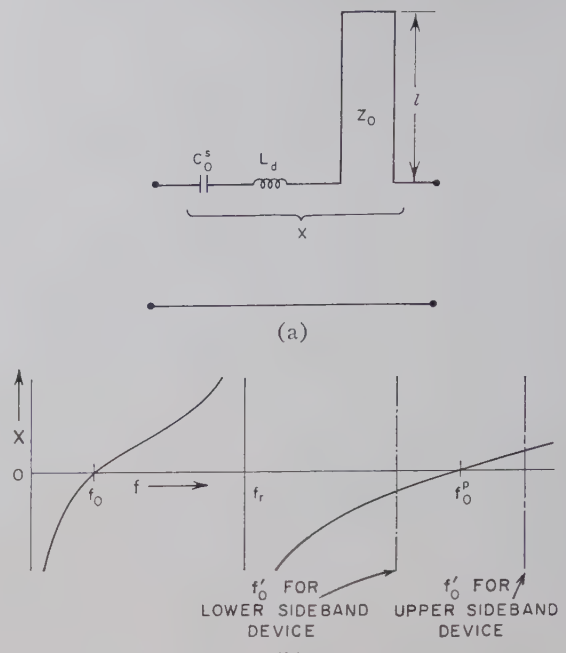


Fig. 6—(a) Approximate equivalent circuit for the diode resonator. (b) Reactance properties of the resonator in Fig. 4.

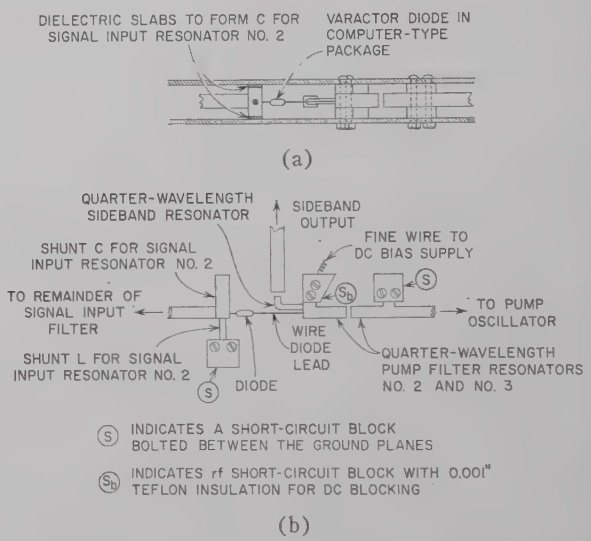


Fig. 7—A possible strip-transmission-line embodiment of the circuit in Fig. 4. The input, pump, and sideband filters are realized in practical microwave filter structures which will appear from the diode resonator circuit to be equivalent to those in Fig. 4 for the frequencies of interest. (a) Cutaway side view. (b) Top view with cover plate removed.

filter fractional bandwidth will be relatively small. Pump filter resonators Nos. 2 and 3 are realized as quarter-wavelength direct-coupled, two-port resonators, since this construction is convenient for narrow or moderate bandwidths, and will not have any spurious pass bands until about three times the frequency of the first pass band.[7] Each resonator bar is series-capacitance-

[6] G. L. Matthaei, "Design of wide-band (and narrow-band) band-pass microwave filters on the insertion loss basis," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MIT-8, pp. 580–593; November, 1960.

[7] G. L. Matthaei, "Direct-coupled, band-pass filters with $\lambda_0/4$ resonators," 1958 IRE NATIONAL CONVENTION RECORD, pt. 1, pp. 98–111.

coupled at one end and shunt-inductance-coupled at the other. Looking to the right from the high Z_0 wire in the diode circuit, the impedance of the structure appears much like Z_e^p seen looking to the right from the high Z_0 line in Fig. 4 for frequencies in the vicinity of f_0^p.

The sideband-output resonator shown in Fig. 7 is one-quarter-wavelength long; it is inductively coupled to the diode and high Z_0 wire circuit, but series-capacitance-coupled to the sideband-output line.

The circuit shown in Fig. 7 appears to be a possible realization of an electronically-tunable up-converter having input frequencies centered around 1000 Mc. However, other configurations are also possible.

V. Determination of Signal-Input and Sideband-Output Circuit Parameters for Upper-Sideband Up-Converters

When power is taken out at the upper sideband while the other components of the mixing process are suppressed, then the $X_{12}X_{21}$ coupling box in Fig. 5 operates like an impedance inverter so that[1]

$$Z_2 = \frac{X_{12}X_{21}}{Z_a'} \tag{17}$$

and

$$Z_2' = \frac{X_{12}X_{21}}{Z_a}, \tag{18}$$

where, as indicated in Fig. 5, the unprimed impedances are evaluated at the signal-input frequency f, while the primed impedances are evaluated at the sideband frequency f'. By the Manley-Rowe equations,[2] the net power P_2 *entering* the left side of $X_{12}X_{21}$ is related to the net power P_2' *entering* the right side of $X_{12}X_{21}$ by

$$\frac{P_2'}{P_2} = -\frac{f'}{f}, \tag{19}$$

where the minus sign indicates that if power flows into the left side, power will flow out of the right side.

Let us define $(R_b)_{\text{nom}}$ as the nominal value of the signal-input filter impedance Z_b in Fig. 5. The impedance $(R_b)_{\text{nom}}$ is purely real and is equal to the resistive terminating impedance which would give best transmission through the fiter. In terms of Fig.3,

$$(R_b)_{\text{nom}} = R_g = \frac{r_g}{r_L} R_{b0} \tag{20}$$

where r_g and r_L are from the lumped-element prototype. For purposes of computing the nominal gain of the amplifier, the signal-input filter will be replaced by a Thevenin equivalent generator having an internal impedance equal to $(R_b)_{\text{nom}}$. At the sideband-output frequency f_0', the impedance Z_b' in Fig. 5 will be purely real and will be defined as

$$R_{b0}' = Z_b'\big|_{f'=f_0'}. \tag{21}$$

Using (17)–(19) and the above definitions for $(R_b)_{\text{nom}}$ and R_{b0}', it can easily be shown that the nominal power gain for input signals in the vicinity of f_0 is

$$\frac{\dfrac{P'_{\text{out}}}{P_{\text{avail}}}}{\dfrac{f_0'}{f_0}} = \frac{4(R_b)_{\text{nom}}R_{b0}'(X_{12}X_{21})_0}{\{[(R_b)_{\text{nom}} + R_s](R_{b0}' + R_s') + (X_{12}X_{21})_0\}^2}, \tag{22}$$

where $R_s = R_s'$ is the diode loss resistance, P_{avail} is the available power at the input frequency f_0, and P_{out}' is the nominal output power at the sideband frequency f_0'. The gain will vary somewhat across the tuning range due to variations in Z_b, $X_{12}X_{21}$, and f'/f, but (22) gives what will herein be referred to as the *nominal gain* of an upper-sideband up-converter.

In the design of an amplifier, the input and pump channel filters serve primarily as wide-band impedance-matching networks. Thus, the first step in the design process is to design the diode resonator circuit and then determine the slope parameters x_1 and x_1^p at frequencies f_0 and f_0^p, respectively, by use of the derivative form in Fig. 3(b). By Fig. 3(b), for a desired fractional bandwidth w of the input circuit, R_{b0} must be

$$R_{b0} = x_1 w \left(\frac{r_L}{\Omega_1 l_1}\right), \tag{23}$$

and then $(R_b)_{\text{nom}}$ is given by (20). The slope parameters x_j and b_k for the other resonators of the input filter may then be computed from the lumped-element prototype filter parameters, R_{b0}, and w, by use of Fig. 3(b) and (e).

Knowing $(R_b)_{\text{nom}}$ and f_0'/f_0, and having values for $(X_{11})_0$, $R_s = R_s'$, and $a = C_1/C_0$ for the diode, the maximum gain will be obtained if R_{b0}' matches Z_2' in Fig. 5, which requires that

$$R_{b0}' = R_s' + \frac{(X_{12}X_{21})_0}{(R_b)_{\text{nom}} + R_s} \tag{24}$$

be satisfied. This can be expressed as

$$R_{b0}' = R_s'\left[1 + \frac{1}{v^2(T+1)}\right], \tag{25}$$

where

$$T = \frac{(R_b)_{\text{nom}}}{R_s} = \frac{(R_b)_{\text{nom}}Q_d}{(X_{11})_0} \tag{26}$$

and

$$v = \frac{1}{aQ_d}\sqrt{\frac{f_0'}{f_0}}. \tag{27}$$

Inserting (25) in (22) yields, after some manipulation,

$$\frac{\dfrac{P'_{\text{out}}}{P_{\text{avail}}}}{\dfrac{f_0'}{f_0}} = \frac{T}{(T+1)[v^2(T+1)+1]}. \tag{28}$$

If this analysis is carried out on the dual basis for the case of the diode resonated in shunt, we find the dual expressions

$$G_{b0} = b_1 w \left(\frac{r_L}{\Omega_1 l_1} \right), \tag{29}$$

$$(G_b)_{\text{nom}} = G_g = \frac{r_g}{r_L} G_{b0}, \tag{30}$$

$$G_{b0}' = G_d' \left[1 + \frac{1}{v^2(T+1)} \right], \tag{31}$$

$$T = \frac{(G_b)_{\text{nom}}}{G_d} = \frac{(G_b)_{\text{nom}} Q_d}{(B_{11})_0}, \tag{32}$$

and (27) and (28) apply as before. The parameter G_d' in (31) is again the equivalent shunt diode loss conductance given by (9). In order to facilitate design calculations, a plot of $(P_{\text{out}}'/P_{\text{avail}})/(f_0'/f_0)$ vs v is shown in Fig. 8 for various values of T. The reference form for the band-pass filters is the dual of that in Fig. 3; the slope parameters and terminations for this dual filter may be conveniently specified directly in terms of the low-pass prototype in Fig. 1.

VI. Pump Circuit Impedance-Matching Filter

In order that the pump channel have a nearly uniform reflection loss across the pump frequency band, it is necessary to design an impedance-matching filter as shown in Fig. 9. In this case, the resistance R_g^p is given by (16b). The No. 1 resonator, having a reactance X_1^p [given by (16b) for the case of Figs. 4, 7 and 9] is characterized by the slope parameter x_1^p computed at frequency f_0^p in Fig. 6(b) using the derivative form in Fig. 3(b). (Recall that in this case, the resonance of the No. 1 resonator is created by the second series resonance of the diode circuit, as discussed in Section IV.) The fractional bandwidth w_p of the pump filter is

$$w_p = w \frac{f_0}{f_0^p}, \tag{33}$$

where w is again the desired fractional bandwidth of the input circuit. By Fig. 3(b),

$$\left(\frac{\Omega_1 l_1}{r_L} \right)_p = \frac{x_1^p w_p}{R_{b0}^p}, \tag{34}$$

where the subscript p on $(\Omega_1 l_1/r_L)_p$ is introduced to indicate that these parameters refer to the low-pass prototype for design of the pump filter. Since, in accord with Figs. 1 and 3,

$$\frac{R_{b0}^p}{R_g^p} = \left(\frac{r_L}{r_g} \right)_p, \tag{35}$$

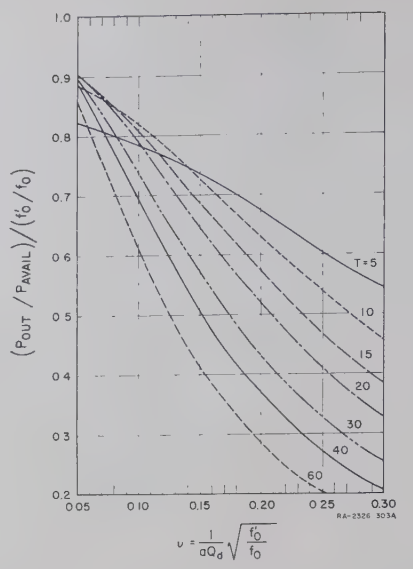


Fig. 8—Chart for determining the gain of upper-sideband up-converters. [$T = (R_b)_{\text{nom}}/R_s$ for the case of a diode resonated in series, and $T = (G_b)_{\text{nom}}/G_d$ for the case of a diode resonated in shunt. This chart does not apply accurately for the shunt resonance case if the diode has significant series inductive reactance.]

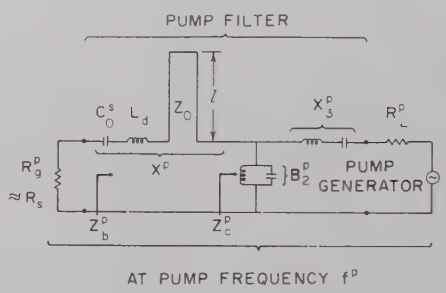


Fig. 9—A circuit approximately equivalent to that in Fig. 4 for energy at the pump frequency f^p.

we may write (34) as

$$\left(\frac{\Omega_1 l_1}{r_g} \right)_p = \frac{x_1^p w_p}{R_g^p}, \tag{36}$$

where, for the case of Figs. 4, 7 and 9, R_g^p is given by (16b). The problem now focuses on finding a low-pass prototype having $(\Omega_1 l_1/r_g)_p$ as given by (36), with a desired amplitude of Tchebycheff pass-band attenuation ripple, and with a minimum amount of reflection loss in the pass band. We are stuck with the prototype impedance $j\Omega l_1 + r_g$, and the other elements of the prototype filter are introduced to optimize the impedance across the band $\Omega = 0$ to $\Omega = \Omega_1$.

Certain aspects of impedance-matching problems of this type were treated by Bode.[8] Fano treated the general limitations on lossless impedance matching, and also the design of certain Tchebycheff impedance-

[8] H. W. Bode, "Network Analysis and Feedback Amplifier Design," D. Van Nostrand Co., Inc., New York, N. Y., pp. 363–368; 1945.

matching networks,[9] and Carlin,[10] Green,[11] Barton,[12] Matthaei,[13] and others have made various contributions.

The problem at hand can be treated as follows. Suppose that it is desired to use n resonators in the pump-circuit filter and that a Tchebycheff ripple of T_r db is specified. Compute

$$H = \text{antilog}_{10}\frac{T_r}{10}, \tag{37}$$

$$d = \sinh\left[\frac{\sinh^{-1}\sqrt{\frac{1}{H-1}}}{n}\right], \tag{38}$$

and

$$e = d - 2\left(\frac{r_g}{\Omega_1 l_1}\right)\sin\frac{\pi}{2n}. \tag{39}$$

Then, for given $r_g/(\Omega_1 l_1)$, T_r, and n, the reflection coefficient value at the maximum reflection points of the Tchebycheff pass band is given by

$$|\rho|_{\max} = \frac{U_n(e)}{U_n(d)}. \tag{40}$$

where $U_n(e)$ and $U_n(d)$ are polynomials tabulated by Matthaei[13] and plotted in Fig. 10 of this paper. Under these conditions, the minimum pump power delivered to the diode in the pump filter pass band is

$$A = \frac{P_{\text{transmitted}}}{(P_{\text{avail}})^p} = 1 - |\rho|^2_{\max} \tag{41}$$

times the available power $(P_{\text{avail}})^p$ of the pump signal generator. It will be found that the larger the slope parameter x_1^p and the larger the fractional bandwidth w_p, the smaller A must be and the more available power will be required.

Having values for $r_g/(\Omega_1 l_1)$, T_r, and $|\rho|_{\max}$ for the pump filter, the actual element values of the low-pass prototype are in this case obtained most easily by use of equations given by Green.[11,14] It should be understood that (37)–(41) apply only to simple ladder circuits such as that in Fig. 1 or their band-pass equivalents (such as that in Fig. 9).

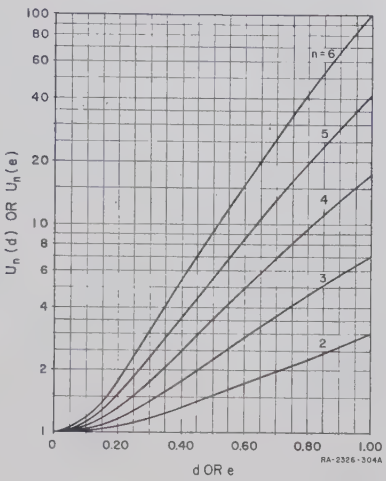


Fig. 10—Chart for use in determining the reflection loss required in the pump circuit when using an optimum Tchebycheff impedance-matching filter.

VII. Determination of Signal-Input and Sideband-Output Circuit Parameters for Lower-Sideband Up-Converters

If power is taken out at the lower sideband instead of at the upper sideband, (17)–(19) become, respectively,

$$Z_2 = -\frac{X_{12}X_{21}}{(Z_a')^*}, \tag{42}$$

$$Z_2' = -\frac{X_{12}X_{21}}{(Z_a)^*}, \tag{43}$$

and

$$\frac{P_2'}{P_2} = \frac{f'}{f}, \tag{44}$$

where the asterisk indicates the complex conjugate. Because of the negative sign in (42), this type of operation gives gain at the input frequency f, due to negative input resistance. As is implied by (44), the power P_2' (at the lower-sideband frequency f') going out the right side of $X_{12}X_{21}$ in Fig. 5 is larger than the power P_2 at frequency f going out the left side by the factor f'/f. Thus, whatever gain is achieved at the input frequency is increased at the $X_{12}X_{21}$ box by the factor f'/f for power taken out at the sideband frequency f'.

In this case, (20) and (21) apply as before, but (22) becomes

$$\frac{P'_{\text{out}}}{P_{\text{avail}}} = \left(\frac{f_0'}{f_0}\right)\frac{4(R_b)_{\text{nom}}R_{b0}'(X_{12}X_{21})_0}{\{[(R_b)_{\text{nom}}+R_s](R_{b0}'+R_s') - (X_{12}X_{21})_0\}^2}, \tag{45}$$

[9] R. M. Fano, "Theoretical limitations on the broadband matching of arbitrary impedances," *J. Franklin Inst.*, vol. 249, pp. 57–83, 139–154; January and February, 1950.

[10] H. J. Carlin, "Synthesis techniques for gain-bandwidth optimization in passive transducers," Proc. IRE, vol. 48, pp. 1705–1714; October, 1960.

[11] E. Green, "Amplitude-Frequency Characteristics of Ladder Networks," Marconi's Wireless Telegraph Co., Ltd., Chelmsford, Essex, England, pp. 62–78; 1954.

[12] B. F. Barton, "Design of Efficient Coupling Networks," Electronic Defense Group, University of Michigan, Ann Arbor, Tech. Rept. No. 44, Contract DA 36-039 SC-63203; March, 1955.

[13] G. L. Matthaei, "Synthesis of Tchebycheff impedance-matching networks, filters, and interstages," IRE Trans. on Circuit Theory, vol. CT-3, pp. 162–172; September, 1956.

[14] E. Green, "Synthesis of ladder networks to give Butterworth or Chebyshev response in the pass band," *Proc. IEE*, pt. 4, Monograph no. 88; 1954.

which gives the nominal gain of the amplifier. The value of R_{b0} for the input filter is determined by (23) as before, but R_{b0}' cannot be determined by impedance matching to maximize the gain as was done to obtain (24), since the maximum possible gain is now infinite (yielding oscillation). In this case, we define a stability parameter

$$D = \frac{(R_b)_{\text{nom}} + R_s}{|Z_2|_{f=f_0}} = \frac{(R_b)_{\text{nom}} + R_s}{\left[\frac{(X_{12}X_{21})_0}{R_{b0}' + R_s'}\right]}, \tag{46}$$

which is seen to be the ratio of the nominal positive resistance of Z_a in Fig. 5 to the magnitude of the negative resistance Z_2 seen looking in the opposite direction. Parameter D fixes the nominal value of the negative resistance gain of the amplifier, and for stability, D must be greater than one. By (46)

$$R_{b0}' = \frac{D(X_{12}X_{21})_0}{(R_{b0} + R_s)} - R_s' \tag{47}$$

$$= R_s'\left[\frac{D}{v^2(T+1)} - 1\right], \tag{48}$$

where T and v have the same meaning as in (26) and (27), respectively. Substituting (48) in (45) gives

$$\frac{\frac{P'_{\text{out}}}{P_{\text{avail}}}}{\frac{f_0'}{f_0}} = \frac{4\left[D - \left(v^2T + \frac{D}{T+1}\right)\right]}{(D-1)^2}. \tag{49}$$

The ratio $(P_{\text{out}}'/P_{\text{avail}})/(f_0'/f_0)$ is plotted vs v in Fig. 11(a)-(c) for $D=2$, 3, 4 and various T values. These charts may also be used for the dual case where the diode is resonated in shunt by defining T as in (32). For the dual shunt-diode case, (48) becomes the dual equation

$$G_{b0}' = G_d'\left[\frac{D}{v^2(T+1)} - 1\right]. \tag{50}$$

VIII. Estimated Performance of Some Design Examples

Let us estimate the performance of some up-converters having an input band center of $f_0 = 1$ kMc. The construction in Figs. 4 and 7 will be assumed using a Hughes 1N896 diode which has the computer-type package shown in Fig. 7. The lead wires on the diode will serve as the high Z_0 line. The wires are 0.020 inch in diameter, and if a 0.500-inch ground-plane spacing is used, they will have an impedance of 207 ohms. The diode will be assumed to have a C_0 value of 1.2 $\mu\mu$f (corresponding to a zero bias capacitance of about 2.3 $\mu\mu$f). The series inductance L_d of the diode itself will be taken as 4×10^{-9} h, and the operating Q will be taken as

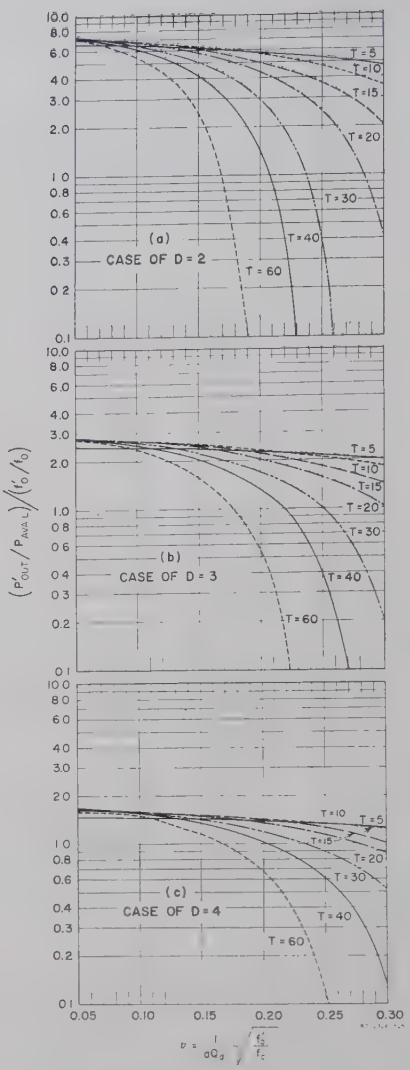


Fig. 11—Charts for determining the gain of lower-sideband up-converters. [$T = (R_b)_{\text{nom}}/R_s$ for the case of a diode resonated in series, and $T = (G_b)_{\text{nom}}/G_d$ for the case of a diode resonated in shunt. Parameter D is the ratio of the magnitude of positive resistance to negative resistance seen at the diode (or for the shunt case, positive conductance to negative conductance). These charts do not apply accurately for the shunt resonance case if the diode has significant series inductive reactance.]

$Q_d = 35$ at $f_0 = 1$ kMc (from the manufacturer's data, this appears to correspond to a cutoff frequency of around 70 kMc for this type of abrupt junction diode). The $a = C_1/C_0$ parameter of the diode will be taken to be 0.25.

In terms of the approximate equivalent circuit in Fig. 6(a), for the diode resonator,

$$X(f) = \frac{-1}{2\pi f C_0^s} + 2\pi f L_d + Z_0 \tan\left(\frac{\pi f}{2f_r}\right), \tag{51}$$

where f_r is the frequency shown in Fig. 6(b) for which the high Z_0 line is $\lambda/4$ long. Setting $f = f_0 = 10^9$, $C_0^s = C_0(1-a^2) = 1.17\times10^{-12}$fd, $L_d = 4\times10^{-9}$h, and $Z_0 = 207$, we find that $f_r = 3.21$ kMc is required, which in turn calls for the $Z_0 = 207$-ohm line to be $l = 0.922$ inch

long. The slope parameter x_1 at frequency f_0 is obtained from

$$x_1 = \frac{\omega_0}{2}\frac{dX}{d\omega}\bigg|_{\omega=2\pi f_0=\omega_0}$$
$$= \frac{1}{2}\left[\frac{1}{2\pi f_0 C_0^s} + 2\pi f_0 L_d + \left(\frac{\pi f_0}{2f_r}\right)\frac{Z_0}{\cos^2\left(\frac{\pi f_0}{2f_r}\right)}\right], \quad (52)$$

which gives $x_1 = 145$ ohms. Solution of (51) shows that f_0^p in Fig. 6(b) will occur at about $f_0^p = 5.45$ kMc, and by calculations similar to (52), the slope parameter there is about $x_1^p = 430$ ohms.[15]

Let us first consider an upper-sideband up-converter having an input tuning band-edge ratio of one-half octave (*i.e.*, for the input filter, $f_2/f_1 = \sqrt{2}$ where f_1 and f_2 are defined in Fig. 2). This calls for $w = 0.348$. The input filter will be assumed to have $n = 3$ resonators and 1-db Tchebycheff ripple. From tables [4,5] of low-pass prototype filter elements, we find that $(\Omega_1 l_1/r_g) = 2.024$ and $r_g = r_L$ for this case. Then by (20) and (23), we obtain $(R_b)_{\text{nom}} = R_{b0} = 25$ ohms. Since $R_s = (X_{11})_0/Q_b = 3.86$ ohms, $T = (R_b)_{\text{nom}}/R_s = 6.5$. In this upper-sideband case, $f_0' = f_0^p + f_0 = 5.45 + 1 = 6.45$ kMc which gives, by (27), $v = 0.291$. By use of T and v in Fig. 8, along with the fact that $f_0'/f_0 = 6.45$, we obtain $P'_{\text{out}}/P_{\text{avail}} = 3.48$ or a nominal gain of 5.3 db. By (25), it is found that the sideband resonator should be adjusted to couple a resistance of $R_{b0}' = 9.95$ ohms into the diode circuit at the upper-sideband frequency $f_0' = 6.45$ kMc.

Using $x_1^p = 430$ ohms, $w_p = wf_0/f_0^p = 0.0638$, and $R_g^p = R_s = 3.86$ ohms in (36), we get $(\Omega_1 l_1/r_g)_p = 7.1$ for the pump filter low-pass prototype. A Tchebycheff ripple of 0.5 db using an $n = 3$ resonator pump filter will be specified which by (38) calls for $d = 0.6265$. Then by (39), $e = 0.4855$. By Fig. 10, $U_3(0.6265) = 3.05$ and $U_3(0.4855) = 2.18$. By (40), $|\rho|_{\max} = 0.715$. By (41), a minimum power of $A(P_{\text{avail}})^p = 0.49\ (P_{\text{avail}})^p$ will be delivered to the diode where $(P_{\text{avail}})^p$ is the available pump power. Assuming that the diode is to be pumped with a peak-to-peak voltage of 6 v, it is estimated that the diode will absorb about 30 mw of pump power when $f^p = f_0^p = 5.45$ kMc. Thus, the required available pump power will be about $(P_{\text{avail}})^p = 30/A = 30/0.49 = 61$ mw, neglecting dissipation loss in the pump input filter. A 1-db dissipation loss would raise this figure to about 77 mw required available power.

Let us now make analogous approximate calculations as described in Section VII using the same diode and diode circuit having $f_0 = 1$ kMc, $f_0^p = 5.45$ kMc, $x_1 = 145$ ohms, $x_1^p = 430$ ohms, etc., as before. Again, $w = 0.348$ will be used (half-octave input tuning range). In this case, the output will be taken at the lower-sideband frequency: $f_0' = f_0^p - f_0 = 5.45 - 1 = 4.45$ kMc. Since this gives negative-resistance operation which increases the sensitivity to input impedance variations, the $n = 3$ resonator input filter will be assumed to have its Tchebycheff ripple reduced to 0.25 db which calls for $(\Omega_1 l_1/r_g) = 1.3034$.

The calculations as outlined in Section VII call for $(R_b)_{\text{nom}} = R_{b0} = 38.7$ ohms, $R_{b0}' = 8.0$ ohms (when $D = 2$). Assuming a stability parameter of $D = 2$, the nominal power gain would be 13.4 db, 8.1 db greater than for corresponding upper-sideband design. The pump power required should be the same in both cases.

Similar approximate calculations for a one-octave-tuning-range, lower-sideband up-converter were made, again assuming the same diode circuit and the same $f_0, f_0^p, f_0' = 4.45$ kMc values. A value of 0.25-db Tchebycheff ripple was assumed for the $n = 3$ resonator input filter. Since stability might be more of a problem using an octave input bandwidth, $D = 3$ was used. In this case, $(R_b)_{\text{nom}} = 78.8$, $R_{b0}' = 5.45$, the nominal gain is 8.8 db, $A = 0.273$, and the required available pump power is about $(P_{\text{avail}})^p = 30/0.273 = 110$ mw, neglecting pump filter dissipation loss and assuming that the diode needs 30 mw of absorbed pump power. If the pump filter has 1-db dissipation loss, the required available power would be about 138 mw.

IX. Conclusions

Properly designed devices of this type using voltage-tunable pump oscillators should have tuning ranges of the order of a half-octave to an octave, fast tuning capability, a useful amount of gain, no image response, and a low noise figure. Though some reflection loss in the pump circuit will be unavoidable, it appears that in many cases existing voltage-tunable oscillators will have adequate incident power.

[15] In computing R_g^p and x_1^p, the denominator in (16b) was neglected and the representation in Fig. 6(a) was used. This should cause little error in the performance estimates about to be computed since, as has been mentioned in Section IV, the denominator of (16b) will be varying quite slowly. It can therefore be regarded as constant within the frequency range of interest. As a constant, it would decrease the size of both R_g^p and x_1^p proportionally (about 25 per cent in this case), but would cancel out in the computation given by (36).

Anisotropic Properties of Strip-Type Artificial Dielectric*

N. J. KOLETTIS† AND R. E. COLLIN‡, SENIOR MEMBER, IRE

Summary—Theoretical formulas for the propagation phase constant of a two-dimensional strip medium are presented for general directions of propagation. In addition a number of experimental results are included that verify the validity of the theory. Some of the difficulties encountered in defining equivalent dielectric constants for this medium are also pointed out.

ARTIFICIAL dielectric media, originally introduced by Kock,[1] have received extensive investigation during the past decade. However, for the most part the detailed study of the anisotropic properties of typical structures has not been carried out. Most studies have been restricted to angles of propagation through the medium and to polarizations, so that the artificial dielectric behaved essentially as an isotropic medium. Since the anisotropic properties can be useful in practice (for example, in wave polarizers), it is worthwhile to analyze these properties for typical structures. From another point of view, anisotropic properties are usually present and hence a knowledge of these properties is needed in order to determine the actual behavior of an artificial dielectric medium in a given application.

The two-dimensional strip-type medium illustrated in Fig. 1 is a typical example of an artificial dielectric that exhibits pronounced anisotropic properties. For propagation through the medium in a direction normal to the broad side of the strips (along x axis) and with the electric field directed along the z axis, the medium behaves as an isotropic phase-delay dielectric. For the opposite polarization the medium behaves as a phase-advance dielectric. The properties of this medium with propagation restricted to be along the x-axis, has been analyzed by Brown,[2] and Cohn,[3] as well as others. If a wave with the electric field polarized in the xz plane is permitted to propagate through the medium at an effective angle θ_r' with respect to the x axis, then the medium exhibits electric anisotropic properties. The analysis for this more general case has been presented by Collin.[4] The purpose of the present paper is to give a summary of the theroetical results obtained to date, together with experimental results that show that the theoretical formulas specifying the anisotropic behavior of the medium give results of acceptable accuracy for design purposes.

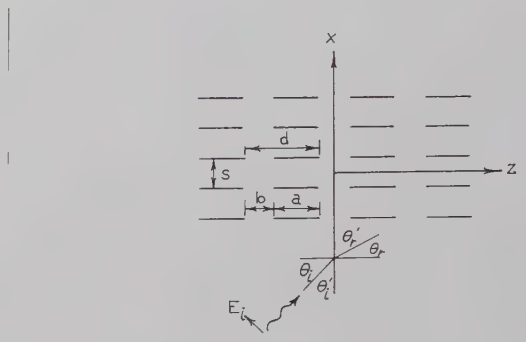


Fig. 1—Cross-section view of two-dimensional strip-type artificial dielectric, strips infinite in length along y-axis.

Modes of Propagation

Before presenting the theoretical formulas that permit the equivalent dielectric constant of the medium to be evaluated, a qualitative discussion of the possible modes of propagation in the medium will be given. If propagation is restricted to be in the x direction with the E vector along the z axis then the dominant mode is a TEM mode and all higher order modes (assumed evanescent) are TM modes with respect to the x axis. These may be labeled as $(TM)_x$ modes. If propagation is permitted to be at an angle to the x axis but in the xz plane and furthermore with E in the xz plane then all the modes are $(TM)_x$ modes. For the opposite polarization with the H vector in the xz plane all the modes are $(TE)_x$ modes. For more general directions of propagation through the medium, so that the propagation vector has a component along all three axes, the possible modes are not of the above type. One set of possible modes of propagation is still characterized by the absence of a y component of electric field, but it does have an x component of H, and hence these may be labeled as $(TE)_y$ modes, *i.e.*, transverse electric modes with respect to the y axis. These are the same types of modes that occur in connection with capacitive diaphragms in rectangular waveguides.[5] These modes may be expressed in terms of the field component H_y. Between

* Received by the PGMTT, April 13, 1961; revised manuscript received, June 30, 1961. This work was sponsored by the Electronics Res. Directorate of the AF Cambridge Res. Lab. under Contract No. AF 19(604)-3887.

† RCA Lab., Princeton, N. J. Formerly with the Case Institute of Technology, Cleveland, Ohio.

‡ Elec. Engrg. Dept., Case Institute of Technology, Cleveland, Ohio.

[1] W. E. Kock, "Metallic delay lenses," *Bell Syst. Tech. J.*, vol. 27, pp. 58–82; January, 1958.

[2] J. Brown, "The design of metallic delay dielectrics," *Proc. IEE*, vol. 97, pt. 3, pp. 45–48; January, 1960.

[3] S. B. Cohn, "Analysis of the metal strip delay structure for microwave lenses," *J. Appl. Phys.*, vol. 20, pp. 257–262; March, 1949.

[4] R. E. Collin, "Field Theory of Guided Waves," McGraw-Hill Book Co., Inc., New York, N. Y. Sec. 12.6; 1960.

[5] H. M. Altschuler and L. O. Goldstone, "On network representations of certain obstacles in waveguide regions," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-7, pp. 213–221; April, 1959.

any two strips the dominant mode in this class is of the form, with the z axis considered as the axis of propagation,

$$H_y = e^{-jk_0(n_y y + n_z z)},$$

where n_y, n_z are components of the unit wave normal $\boldsymbol{n}$. In addition there are field components H_z and E_x. This mode has no low-frequency cutoff.

The other class of possible modes is characterized by the absence of a y component of magnetic field and may be classed as $(TM)_y$ modes. These modes may be expressed in terms of E_y. The dominant mode in this class that may exist between two strips, again considering the z axis as the axis of propagation, is of the form

$$E_y = \sin\frac{\pi x}{s} e^{-jk_0(n_y y + n_z z)}$$

where $(k_0 n_z)^2 = k_0^2(1 - n_y^2) - (\pi/s)^2$. For the usual values of s occurring in practice, $(s < \lambda_0/2)$, n_z is imaginary and this mode is below cutoff. Thus for the $(TM)_y$ modes the strip-medium has a low frequency stop band. Only for $s > \lambda_0/2$ can a propagating $(TM)_y$ mode exist, and when it does the medium functions as a phase-advance medium for these modes.

In this paper we are concerned only with the case of the $(TE)_y$ modes. For these, the component of the electric field along the z axis gives rise to a large z-directed electric dipole moment in each strip and hence an effective dielectric constant κ_e in the z direction. The strips are infinitely thin in the x direction; hence E_x produces no dipole polarization in the x direction, so the effective dielectric constant in this direction is unity. The dielectric constant in the y direction may be arbitrarily chosen since the $(TE)_y$ modes do not have an electric field component in this direction. For reasons to be explained later, it is convenient to chose the effective dielectric constant in the y direction as κ_e.

The specification of equivalent dielectric constants for an artificial dielectric medium is just a convenient artifice that will permit an analogy to be made between the artificial medium and a homogeneous anisotropic dielectric medium. Such an analogy is useful only if the propagation of a wave through the homogeneous dielectric follows the same law as that for the artificial dielectric. When this is true the refractive properties of the artificial medium are the same as for the homogeneous dielectric and the design of a microwave lens may then be based on the properties of the equivalent homogeneous dielectric. It should be noted at this time that a homogeneous medium which is fully equivalent to the artificial dielectric medium for all possible modes of propagation can usually not be specified. Thus, whatever analogy can be established is usually valid only under certain restrictive conditions on wave polarization and direction of propagation.

The homogeneous anisotropic dielectric medium chosen as an analogy for the strip-medium, for a $(TE)_y$ mode of propagation in the strip medium, is a uniaxial medium with dielectric constants 1, κ_e, κ_e along the x, y and z axis respectively. The basis for this analogy is that for the strip medium, the low frequency solution for the propagation phase constant β, along the wave normal $\boldsymbol{n}$, satisfies the same eigenvalue equation as for the homogeneous medium. Furthermore, for propagation in the xz plane, the extraordinary wave in the uniaxial medium has the same field components as the $(TE)_y$ mode in the strip medium. The analogy is, however, not a complete one, since for propagation through the medium with $\boldsymbol{n}$ having a y component, the $(TE)_y$ mode has a zero y component of electric field while the extraordinary wave in the homogeneous uniaxial dielectric has a zero component of magnetic field along the optic axis (x axis in the present case).[6] Nevertheless the eigenvalue equation and refractive properties of the two media are the same for the respective modes of propagation that they support.

Theoretical Results for Equivalent Dielectric Constant

The strip medium illustrated in Fig. 1 is a double periodic medium and may be analyzed in terms of modes propagating along the z axis or in terms of modes propagating along the x axis. In the first method, which will be referred to as the z-axis solution, each layer of strips presents a discontinuity at each interface. The reflection and transmission coefficients that characterize each interface are those associated with the interface between a set of semi-infinite parallel plates and free space. The latter problem has been solved by Carlson and Heins.[7] Using their results, it is found that the equivalent circuit of the strip medium for modes propagating in the z direction is that illustrated in Fig. 2.[4] The circuit in Fig. 2 is arrived at by assuming that the strip spacing s is small compared with a wavelength and that neither a nor b are so small that higher-order mode interaction between adjacent interfaces is important. For a TEM mode incident, in the xz plane, on the strip medium at an angle θ_i relative to the z axis and with the electric vector in the xz plane, the propagation phase constant βn_z through the medium is given by

$$\begin{aligned}(1 - \rho^2)\cos\beta n_z d &= \cos\left[k_0(a + b\cos\theta_i) - k_0\frac{s}{\pi}(1 - \cos\theta_i)\ln 4\right] \\ &\quad - \rho^2\cos\left[k_0(a - b\cos\theta_i) - k_0\frac{s}{\pi}(1 + \cos\theta_i)\ln 4\right], \quad (1)\end{aligned}$$

[6] Collin, *op. cit.*, Sec. 3.7.

[7] J. F. Carlson and A. E. Heins, "The reflection of an electromagnetic plane wave by an infinite set of plates," *Quart. Appl. Math.*, vol. 4, pp. 313–329; January, 1957, and vol. 5, pp. 82–88; April, 1947.

where

$$\rho = (1 - \cos\theta_i)/(1 + \cos\theta_i), \qquad k_0 = 2\pi/\lambda_0$$

and the other parameters are give in Fig. 1. This equation results from an analysis of the periodic structure illustrated in Fig. 2. When the frequency is low so that $k_0 d$ approaches zero, the cosine terms in (1) may be approximated according to the relation $\cos x = 1 - x^2/2$. After simplification (1) reduces to

$$(\beta n_z)^2 = k_0^2\left[1 - \left(\frac{b}{d} + \frac{s}{\pi d}\ln 4\right)\sin^2\theta_i\right]. \qquad (2)$$

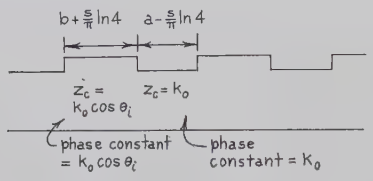


Fig. 2—Equivalent circuit for strip medium for modes propagating in the z direction.

In the present case the propagation phase constant along the x axis is $k_0 \sin\theta_i = \beta n_x$ and is known. If propagation takes place along the y axis also with a phase constant βn_y, then in (1) and (2) k_0^2 must be replaced by $k^2 = k_0^2 - \beta^2 n_y^2$ and βn_x becomes equal to $k \sin\theta_i$. The basis for this substitution is the same as that used to obtain the solution for a capacitive diaphragm in a rectangular guide from the solution for a diaphragm in an infinitely wide parallel-plate transmission line. Thus for general directions of propagation the low frequency limit of (1) is

$$(\beta n_z)^2 = k_0^2 - (\beta n_y)^2 - (\beta n_x)^2\left(\frac{b}{d} + \frac{s}{\pi d}\ln 4\right). \qquad (3)$$

The static solution for the dielectric constant of the strip medium in the z direction, as obtained by conformal mapping, is given by[8]

$$\kappa_e = \frac{d}{b + \dfrac{s}{z}\ln 4} \qquad (4)$$

when both s/a and s/b are small. Substituting this into (3) and solving for β^2 gives

$$\beta^2 = \frac{\kappa_e k_0^2}{n_x^2 + (1 - n_x^2)\kappa_e}. \qquad (5)$$

This, however, is the eigenvalue equation for the extraordinary wave in a homogeneous uniaxial anisotropic dielectric with dielectric constants 1, κ_e, κ_e, along the x, y, z axis, respectively. It is for this reason that the homogeneous uniaxial medium is chosen as an analogy for the strip medium, and also the reason why the strip medium is assumed to have a dielectric constant κ_e in the y direction. Another similar analogy is also possible, and this is discussed in the Appendix.

For high frequencies βn_z must be found from (1). By analogy with (3) the equivalent dielectric constant κ_e may then be defined by

$$(\beta n_z)^2 = k_0^2 - (\beta n_y)^2 - \kappa_e^{-1}(\beta n_x)^2,$$

or

$$\kappa_e = \frac{(\beta n_x)^2}{k_0^2 - (\beta n_z)^2 - (\beta n_y)^2}. \qquad (6)$$

In general, κ_e is a function of frequency except in the low frequency range where (4) is valid.

The solution provided by (1) may be expected to give good accuracy for small values of s. For larger values of s higher mode interaction becomes significant and should be accounted for. On the other hand, by solving the same problem in terms of modes propagating in the x direction, the condition that higher order mode interaction in this approach be negligible, is that the spacing s be large. Thus the x-axis solution is complementary to the z-axis solution in that it will yield good results for large s while the latter yields good accuracy for small s. A combination of the two methods will then enable accurate estimates of κ_e to be obtained for all values of s likely to be encountered.

For the x-axis solution each row of strips in an x = constant plane constitute a capacitive susceptance jB. Thus the equivalent circuit for modes propagating in the x direction is a transmission line loaded at intervals s by shunt susceptances jB. The propagation phase constant βn_x in the x direction is a solution of

$$\cos\beta n_x s = \cos hs - \frac{B}{2}\sin hs, \qquad (7)$$

where

$$h^2 = [k_0^2 - (\beta n_y)^2]\cos^2\theta_i'$$

and θ_i' is the angle of propagation, relative to the x-axis, for the TEM mode between strips when $n_y = 0$. Alternatively h may be expressed as $k_0 \cos\theta_x$ where $\cos\theta_x$ is the direction cosine between the x axis and the wave normal of the TEM wave existing between adjacent gratings, an expression valid for all n_y. In the analysis of the problem βn_y and βn_z are usually specified and hence

$$h^2 = k_0^2 - (\beta n_y)^2 - (\beta n_z)^2, \qquad (8)$$

since for the TEM mode, between gratings, the magni-

[8] J. Howes and E. A. N. Whitehead, "The Refractive Index of a Dielectric Loaded with Thin Metal Strips," Elliot Brothters Ltd., London, England, Elliot Bros. Rept. 119; July, 1949.
N. Kolettis, "Conformal Mapping Solution for Equivalent Relative Permittivity of an Artificial Strip Dielectric Medium," Case Inst. Tech., Cleveland, Ohio, Scientific Rept. No. 2, issued on Contract AF 19(604)-3887; January 15, 1959.

tude of the propagation phase constant must equal k_0. An approximate value of B is given by[9]

$$B = \frac{2k_0 d \cos\theta_i'}{\pi}\left\{ ln \csc\frac{\pi b}{2d} + \frac{1}{2}\,\frac{(1-\alpha^2)^2[(1-\alpha^2/4)(A_+ + A_-) + 4\alpha^2 A_+ A_-]}{(1-\alpha^2/4) + \alpha^2(1+\alpha^2/2-\alpha^4/8)(A_+ + A_-) + 2\alpha^6 A_+ A_-}\right\}, \tag{9}$$

where

$$\alpha = \sin \pi b/2d$$

and

$$A_\pm = \left[1 \pm \frac{k_0 d}{\pi}\sin\theta_i' - \left(\frac{k_0 d\cos\theta_i'}{2\pi}\right)^2\right]^{-1/2} - 1.$$

In this equation, $k_0 \cos\theta_i' = h$ and $k_0 \sin\theta_i' = (k_0^2 - h^2)^{1/2}$ and h is given by (8) in general. When $n_y = 0$, θ_i' is the angle of incidence relative to the x axis in Fig. 1, but for $n_y \neq 0$, k_0^2 is to be replaced by $k^2 = k_0^2 - (\beta n_y)^2$ and $h = k_0 \cos\theta_i'$ becomes $(k_0^2 - \beta^2 n_y^2)^{1/2} \cos\theta_i'$ while $k_0 \sin\theta_i'$ becomes $(k_0^2 - \beta^2 n_y^2)^{1/2} \sin\theta_i' = (k_0^2 - \beta^2 n_y^2 - h^2)^{1/2} = \beta n_z$. Once βn_x has been found from (7) the equivalent dielectric constant κ_e may be obtained from (6) since βn_z and βn_y are determined by the known incident wave on the medium.

Experimental Results

Measurement of the propagation constant βn_z was carried out for six different samples three of which were eight periods long, while the other three were ten periods long. The samples were constructed by mounting metallic foil strips between sheets of polyfoam as illustrated in Fig. 3. Each sample was made to fit inside a rectangular guide of internal dimensions 1 7/8″ by 7/8″ and having a removable top plate with a centered slot. The foil strips were bent over along the sides of each sample in order to make a reasonably good contact with the side walls of the guide. At each end of the sample the guide was closed off by a short-circuiting block so as to form a resonant cavity. The cavity was excited through a small circular aperture in one end plate and with an E_{11} mode incident in the main guide. When the cavity resonates $\beta n_z = n\pi/Nd$ where n is the nth resonant mode and N is the total number of sections in the sample. The available pass band of the main guide for the E_{11} mode permitted from 4 to 7 resonant frequencies to be found. Resonance was detected by means of a probe inserted into the cavity through the centered slot along the top surface of the cavity.

The measured values of βn_z for strips mounted on polyfoam may be converted to equivalent results for strips in free space by frequency scaling. If f_1 is a resonant frequency for the cavity filled with strips mounted on polyfoam, then $\kappa_p^{1/2} f_1$ is the resonant frequency of the same mode for a cavity filled with strips surrounded by free space. κ_p is the dielectric constant of polyfoam, about 1.08 in the experiments carried out. This frequency scaling was used to convert all results to that for strips surrounded by free space, so as to avoid the need to consider the properties of the supporting medium in the comparison of the theoretical and experimental results.

In the z-axis solution for the loaded guide $\beta n_x = \pi/d_1$ and $\beta n_y = \pi/d_2$ where d_1 is the guide height (7/8″) and d_2 is the guide width (1 7/8″). In the x-axis solution βn_x and βn_y are also fixed by the guide dimensions and hence (7) and (9) must be solved simultaneously for the parameter h. Eq. (8) is then readily solved for βn_z. The solution for h was obtained by plotting B vs h as determined by (7) and (9). The point of intersection of the two curves determines the value of h that satisfied both equations.

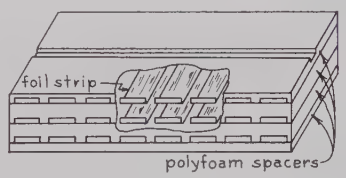


Fig. 3—Typical strip medium sample with 3 layers and 8 sections.

In Figs. 4 and 5 the theoretical and experimental values of βn_z are plotted as a function of frequency. The frequency has been scaled so that the results are for strips surrounded by free space. For Fig. 4 the parameters of the three samples measured are $a = 1.143$ cm, $b = 0.381$ cm, $s = 0.554$ cm, 0.739 cm, 1.11 cm, respectively and $N = 8$, *i.e.*, the samples are 8 sections long. The samples for Fig. 5 are 10 sections long and $a = 0.762$ cm, $b = 0.254$ cm, and $s = 0.554$ cm, 0.739 cm, and 1.11 cm, respectively. An examination of the experimental results shows very clearly that for small values of s the experimental values lie very close to the theoretical curve based on the z-axis solution, while for the largest value of s used the experimental results lie close to the theoretical curve based on the x-axis solution. Fig. 6 is a plot of the computed value of κ_e [based on (6)] as a function of frequency for the two samples for which $s = 0.739$ cm. At the higher frequencies κ_e rises rapidly in value since the edge of the first pass band is being approached.

[9] N. Marcuvitz, "Waveguide Handbook," McGraw-Hill Book Co., Inc., New York, N. Y., vol. 10 of M.I.T. Rad. Lab. Series, Sec. 5.18, Eq. (1a); 1951.

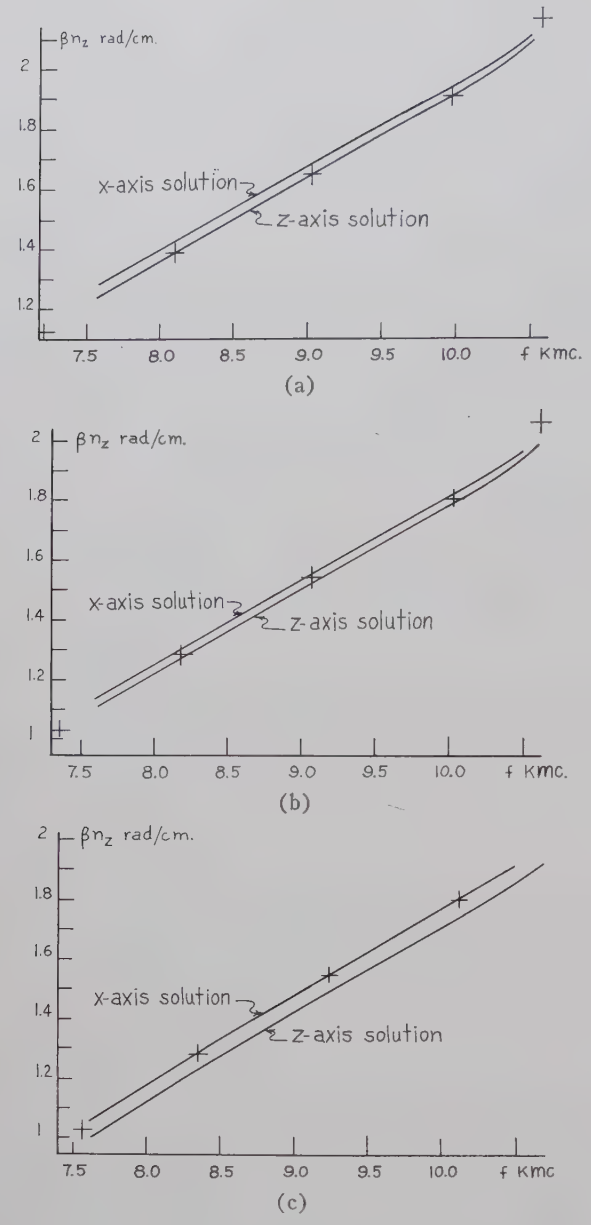


Fig. 4—Computed and measured values of βn_z for $a=1.143$ cm, $b=0.381$ cm, + experimental points. (a) $s=0.554$ cm, (b) $s=0.739$ cm, (c) $s=1.11$ cm.

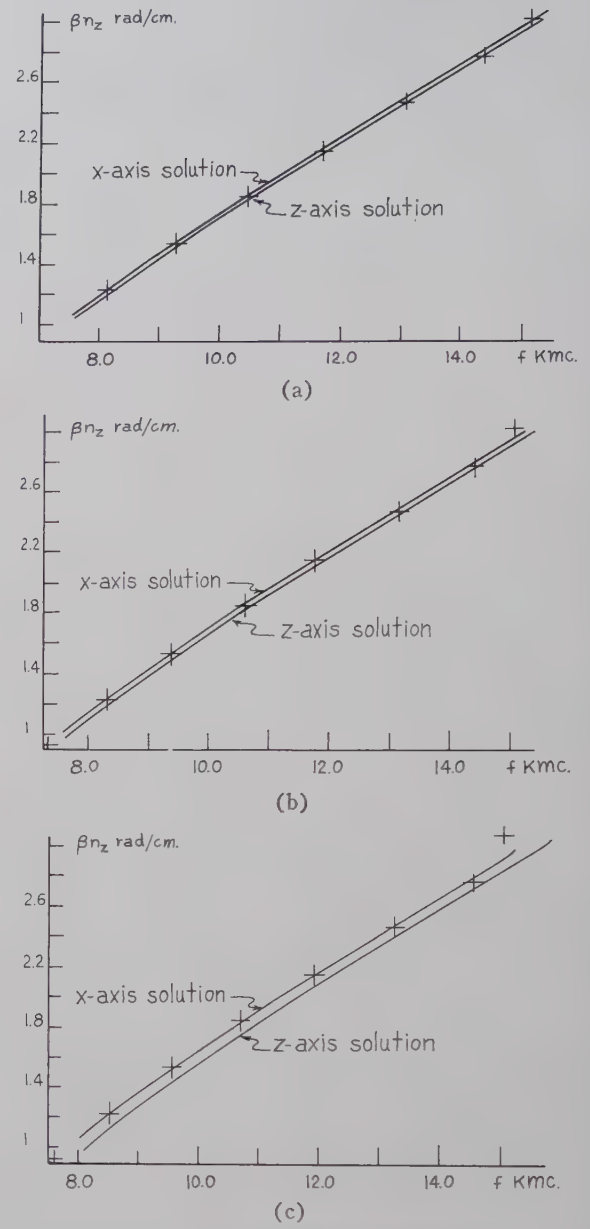


Fig. 5—Computed and measured values of βn_z for $a=0.762$ cm, $b=0.254$ cm, + experimental points. (a) $s=0.554$ cm, (b) $s=0.739$ cm, (c) $s=1.11$ cm.

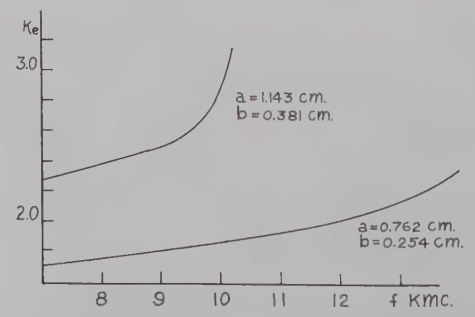


Fig. 6—Computed value of κ_e for two samples for which $s=0.739$ cm, based on experimental values of βn_z from Figs. 4(b) and 5(b).

The theoretical solution for βn_z, in the case $n_y=0$, yields βn_z as an implicit function of the free space wave number k_0. When $n_y \neq 0$ the solution for βn_z at any given frequency f_1 may be found by simply replacing $k_0=2\pi f_1(\mu_0\epsilon_0)^{1/2}$ by $k=(k_0^2-\beta^2 n_y^2)^{1/2}$. Thus the curves presented in Figs. 4, 5, and 6 as a function of frequency with $\beta n_y=\pi/d_2$, may be interpreted as results for the case $n_y=0$, *i.e.*, propagation in xz plane only, by replacing the frequency scale f by a new frequency scale f_1 where

$$f_1=\left[f^2-\frac{1}{4\mu_0\epsilon_0 d_2^2}\right]^{1/2}.$$

Conclusions

Experimental results have been presented which show that for small strip spacing s, the z-axis solution gives good results for the propagation phase constant of the strip-type artificial dielectric medium. For larger values of the spacing s, a solution based on modes propagating normal to the broad face of the strips gives good results. A combination of the two methods thus provides a method of estimating the value of the propagation constant without having to resort to the laborious procedure of taking higher-order mode interaction into account. In order to estimate the best value of the propagation phase constant it is necessary to plot βn_z as a function of the strip spacing s with other parameters fixed. When this is done it is found that for a certain range of s, the two methods give results that are in close agreement. For smaller values of s the best estimate curve of βn_z should be drawn so as to approach the z-axis solution, while for large values of s the best estimate curve for βn_z should be made to approach the x-axis solution. Typical results that are obtained by this procedure are given in Collin.[4]

From a knowledge of the propagation constant βn_z, an equivalent dielectric constant for the medium, along the z and y axis, may be defined and evaluated by considering an analogy between the strip-type artificial dielectric and a homogeneous uniaxial anisotropic dielectric medium. Some of the difficulties encountered in formulating such an analogy were pointed out. In particular, it was shown that a complete analogy with a homogeneous anistropic dielectric medium was not possible in general.

Appendix

The strip-medium could be assumed to have a dielectric constant of unity in both the x and y directions. In this case one would attempt to make an analogy with a uniaxial medium with dielectric constants 1, 1, κ_e along the x, y and z axis. The z-axis now becomes the optic axis. In this case the eigenvalue equation for the homogeneous dielectric medium is

$$\beta^2=\frac{\kappa_e k_0^2}{n_z^2\kappa_e+(1-n_z^2)}.$$

Only if $n_y=0$ so $n_z^2=1-n_x^2$ is this equation the same as (5). Since (5) arises naturally from the low-frequency limit of (1) it seems preferable to use the analogy discussed in the paper.

A further point of interest is that if the analogy discussed in this Appendix is used then in place of (6), the dielectric constant is in general given by

$$\kappa_e=\frac{(\beta n_x)^2+(\beta n_y)^2}{k_0^2-(\beta n_z)^2},$$

which is not the same as that given by (6). The analogy discussed in this Appendix is the same as that used by Kolettis.[10]

Although the two analogies lead to different results for the equivalent dielectric constant κ_e, they are both of equal usefulness in practice since, invariably, the parameter κ_e is used to compute β and, provided the appropriate eigenvalue equation is used, the same value of β obviously results. Clearly, since the strip medium is not fully analogous to a uniaxial homogeneous anisotropic dielectric medium, not too much physical significance can be attached to the defined equivalent dielectric constant in either case.

[10] N. J. Kolettis, "Electric Anisotropic Properties of the Metallic-Strip-Type Periodic Medium," Case Inst. Tech., Cleveland, Ohio, Scientific Rept. No. 17 issued on Contract No. AF 19–(604)-3887; October, 1960.

Optimization of Waveguide Tapers Capable of Multimode Propagation*

C. C. H. TANG†

Summary—By converting Maxwell's equations, the general case of mode conversion in tapered waveguides is treated by matrix formulation in terms of an infinite set of coupled differential equations with nonuniform coupling coefficients and varying phase constants. An "orthogonalization" or "diagonalization" process is introduced through a nonlinear matrix transformation which is a function of taper length. The general matrix solution of the problem is obtained through a perturbation method in the form of an integral equation of the Volterra type, and the integral equation is solved by an iteration method. In view of the difficulties in finding eigenvalues, the problem is then reduced to the two-mode case, and the mode conversion is obtained in an explicit form revealing certain information which characterizes the choice of "mode-conversion distribution function." Optimization is first obtained through proper choice of the mode-converison distribution function. In an attempt to approximate a Tchebycheff mode-conversion response, further optimization is realized by creating "new zeros" and thereby changing the density of the distribution of zeros in the vicinity of the origin of the mode-conversion curve and the nature of the optimization procedure essentially becomes that of synthesis. Through using the optimized distribution function, a total reduction of about 50 per cent in taper length is realized (when compared with the cosine-squared distribution) for the case of 50-db prescribed-mode discrimination in a taper connecting a $\frac{7}{8}$-in ID waveguide to a 2-in ID waveguide operating in the circular electric mode up to 75 kmc.

I. Introduction

A PERUSAL of recent literature shows that the tapered waveguide capable of multimode propagation has been a subject of interest for the past few years, since in long-distance transmission by use of the low-loss circular electric mode (TE_{01}) in circular waveguide, tapers are necessary in several important applications. Tanaka,[1] expanding the field of conical guide in terms of the eigenfunctions of the uniform circular guide and then matching the fields at the junction, obtained a general expression for a mode conversion through the taper. Solymar,[2] using a similar technique with some approximations, gave the design procedure of a one-section conical taper and a multisection conical taper. Savvirykh,[3] combining perturbation theory with the method of W.K.B., treated the field in the tapered waveguide in terms of an eigenfunction series expansion in an "artificial" orthogonal coordinate system. Unger[4] obtained an improved design of the circular waveguide taper with varying cone angle by using two coupled differential equations. The present paper treats the general case of an infinite set of coupled differential equations in matrix formulation, and the general matrix solution of the problem is obtained through solving an integral equation of the Volterra type by an iteration method. A procedure to create "new zeros" in mode-conversion characteristics is introduced in order to "optimize" the taper for "minimum" length.

The presence of a taper in a waveguide inevitably introduces spurious modes. On the assumption that the taper possesses perfect symmetry and its axis is perfectly straight, only TE_{0n} modes will be excited in the tapered region. If the taper axis has a slight curvature, additional TM_{1n} modes will be excited due to the degeneracy between the TE_{0n} and TM_{1n} modes. Our goal in designing the taper is to reduce the spurious modes to a prescribed level in the operating frequency range with a taper length as short as possible.

To describe electromagnetic fields in a perfectly conducting curved waveguide with nonuniform and arbitrary cross section, we have to solve Maxwell's equations with appropriate boundary conditions. Schelkunoff[5] has shown that certain field problems with complicated boundary conditions can be handled more readily by converting Maxwell's equations into generalized telegraphist's equations. To convert Maxwell's equations into telegraphist's equations, we introduce a complete set of orthogonal functions most appropriate to the particular geometry of the problem in question. Fields within the guide are then expanded in a series in terms of these orthogonal functions of a complete set. Subsituting these fields into Maxwell's equations with due care given to the convergence property of the series on the boundary and taking advantage of the particular orthogonality relations of the functions, we obtain the following generalized telegraphist's equations:

$$\frac{dV_m}{dw} = -\sum_{n=1}^{\infty} Z_{mn} I_n + \sum_{n=1}^{\infty} {}^{V}T_{mn} V_n$$

$$\frac{dI_m}{dw} = -\sum_{n=1}^{\infty} Y_{mn} V_n + \sum_{n=1}^{\infty} {}^{I}T_{mn} I_n,$$

$$m = 1, 2, 3, \cdots, \quad (1)$$

* Received by the PGMTT, April 17, 1961; revised manuscript received, June 23, 1961.

† Bell Telephone Labs., Murray Hill, N. J.

[1] K. Tanaka, "Mode Conversion Through the Tapered Section of Circular Waveguide System," presented at the Congrès Internatl. Circuits et Antennas Hyperfrequences, Paris, France; October, 1957.

[2] L. Solymar, "Monotonic multisection tapers for over-moded circular waveguides," *Proc. IEE*, vol. 106, pt. B, suppl. no. 13, pp. 121, 1959.

[3] S. K. Savvirykh, "On the theory of tapered circular waveguides," *Radioteckh. i Elektron*, vol. 4, pp. 972–1002; April, 1959.

[4] H. Unger, "Circular waveguide taper of improved design," *Bell Sys. Tech. J.*, vol. 37, pp. 899–912; July, 1958.

[5] S. A. Schelkunoff, "Conversion of Maxwell's equations into generalized telegraphist's equations," *Bell Sys. Tech. J.*, vol. 34, pp. 995–1403; September, 1955.

where w is the coordinate along the assumed direction of propagation, Z_{mn} is the self or mutual series impedances, Y_{mn} is the self or mutual shunt admittances, ${}^{V}T_{mn}$ is the "voltage-transfer coefficients," and ${}^{I}T_{mn}$ is the "current-transfer coefficients." V_m and I_m are, respectively, the voltages and currents related to the amplitudes of electric and magnetic intensities associated with each particular function. For the circular-waveguide taper, we choose as the complete set the orthogonal modes of a straight-circular waveguide with perfectly conducting walls and filled with a homogeneous dielectric as in Fig. 1. It is obvious that the

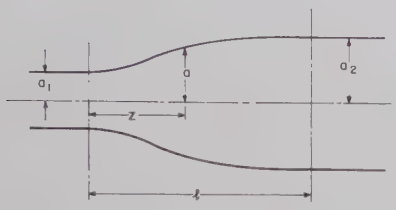


Fig. 1—Circular waveguide taper.

choice is arbitrary, although most appropriate in this case. The infinite set of first-order differential equations (1) with coupling among all possible modes reduces, in this case, to an infinite set with coupling among TE_{0m} modes only and takes the following form[4] in cylindrical coordinates:

$$\frac{dV_m}{dz} = -j\omega\mu I_m + \frac{1}{a}\frac{da}{dz}\sum_{n=1}^{\infty}{}' \frac{2k_m k_n}{k_n^2 - k_m^2} V_n$$

$$\frac{dI_m}{dz} = j\frac{\gamma_m^2}{\omega\mu} V_m + \frac{1}{a}\frac{da}{dz}\sum_{n=1}^{\infty}{}' \frac{2k_m k_n}{k_n^2 - k_m^2} I_n,$$

$$m = 1, 2, 3, \cdots, \quad (2)$$

where the summations are extended over all n except $n=m$. γ_m is the propagation constant of the mth mode, and k_m is the mth root of the Bessel function J_1. If we let A_m and R_m be the amplitudes of the forward and backward waves of a typical mode TE_{0m}, the following equations are always true:

$$V_m = \sqrt{Z_m}\,(A_m + R_m),$$

$$I_m = \frac{1}{\sqrt{Z_m}}(A_m - R_m), \quad (3)$$

where Z_m is the wave impedance of TE_{0m} mode

$$Z_m = \frac{j\omega\mu}{\gamma_m}.$$

Substitution of (3) into (2) results in a new set of equations in A_m and R_m. If the taper is very gradual, we can assume backward waves and multiple reflections negligibly small and obtain the following infinite set of coupled differential equations in forward-wave amplitudes:

$$\frac{dA_m}{dz} = -\gamma_m A_m + \sum_{n=1}^{\infty}{}' \xi_{mn} A_n, \quad (4)$$

where ξ_{mn}'s are the coupling coefficients defined by

$$\xi_{mn} = \frac{1}{a}\frac{da}{dz}\frac{k_m k_n}{k_n^2 - k_m^2}\left(\sqrt{\frac{Z_m}{Z_n}} + \sqrt{\frac{Z_n}{Z_m}}\right). \quad (5)$$

II. General Formulation in Matrix Representation

The infinite set of equations (4) for a section of a tapered waveguide can be conveniently cast in matrix form

$$D\{A\} = [M]\{A\}, \qquad \{A(0)\} = \{C\}, \quad (6)$$

where D, $\{\ \}$ and $[\ ]$ represent, respectively, an operator, a column matrix and a square matrix as follows:

$$D = \frac{d}{dz}, \qquad \{A\} = \begin{bmatrix} A_1(z) \\ A_2(z) \\ A_3(z) \\ \vdots \end{bmatrix},$$

$$[M] = \begin{bmatrix} -\gamma_1(z) & \xi_{12}(z) & \xi_{13}(z) \cdots \\ \xi_{21}(z) & -\gamma_2(z) & \xi_{23}(z) \cdots \\ \xi_{31}(z) & \xi_{32}(z) & -\gamma_3(z) \cdots \\ \vdots & \vdots & \vdots \end{bmatrix}.$$

$[M(z)]$ is assumed to be bounded continuous square-matrix function in a certain interval of the z axis. $\{C\}$ is a column matrix of constants. It is seen that the presence of tapering introduces the coupling terms $\xi_{mn}(z)$ in $[M]$, and the A_m is no longer orthogonal. In a uniform circular waveguide, the coupling coefficients ξ_{mn} vanish, the propagation constants γ_m are independent of z, and the amplitudes A_m form a complete orthogonal set.

For the lossless case, we can show with the aid of the law of conservation of energy that the matrix $[M]$ is, in general, complex and $\xi_{mn} = -\xi_{nm}{}^*$.[6] In our particular case, the elements ξ_{mn} of the matrix are all real, and $\xi_{mn} = -\xi_{nm}$ since the impedances Z_m and Z_n are all real for all modes far away from the cut-off as seen from (5). The propagation constant γ_m is purely imaginary for perfectly conducting tapers and has the form $\gamma = j\beta_m$, where β_m is real. The square matrix $[M]$ then takes the form

$$[M] = \begin{bmatrix} -j\beta_1(z) & C_{12}(z) & C_{13}(z) \cdots \\ -C_{12}(z) & -j\beta_2(z) & C_{23}(z) \cdots \\ -C_{13}(z) & -C_{23}(z) & -j\beta_3(z) \cdots \\ \vdots & \vdots & \vdots \end{bmatrix}, \quad (7)$$

where $\xi_{mn} = C_{mn}$, real functions of z.

[6] The star * implies the complex conjugate.

We now introduce a nonlinear, nonsingular matrix transformation of the following form:

$$\{A\} = [P(z)]\{B(z)\}, \tag{8}$$

in an attempt to obtain a new orthogonal set of amplitudes $B_m(z)$ along the tapered waveguide in at least a localwise sense. Substituting (8) into (6) and following the rules of matrix calculus, we obtain the following:

$$[P]D\{B\} = [M][P]\{B\} - D[P]\{B\}. \tag{9}$$

Multiplication of (9) by $[P]^{-1}$ yields

$$D\{B\} = [Q]\{B\}, \qquad \{B(0)\} = [P(0)]^{-1}\{C\}, \tag{10}$$

where

$$[Q] = [P]^{-1}[M][P] - [P]^{-1}D[P] = [q] - [\epsilon], \tag{11}$$

and

$$[q] = [P]^{-1}[M][P], \qquad [\epsilon] = [P]^{-1}D[P].$$

If the transformation matrix $[P]$ is independent of z, the second term in (11) would vanish and the matrix $[Q]$ is truly diagonal and the B_m's would be orthogonal as seen from (10). In general, the matrix $[Q]$ is not diagonal because of the presence of the term $[P]^{-1}D[P]$, which represents the effect of tapering. However, if $[P]^{-1}D[P]$ can be considered as a small perturbing term to the uniform guide, the matrix differential equation (10) can be solved approximately by perturbation methods. It is worthwhile to note that the transformed-matrix differential equation (10) is of the same form as the original differential equation (6); however, the significance lies in the fact that the weights of the nondiagonal terms of the matrix $[M]$ have been shifted into the diagonal terms of the matrix $[Q]$ by the transformation of (8). Eq. (10), therefore, can yield a more accurate solution by the perturbation method than (6).

To find the matrix $[Q]$, we must first obtain the eigenvectors comprising the diagonalizing matrix $[P]$. This involves the determination of eigenvalues from the secular equation of the matrix

$$|[M] - \lambda[I]| = 0, \tag{12}$$

where $[I]$ is the unit matrix. The values of λ for which the equation is satisfied are the desired eigenvalues. Since $[M]$ is, in general, complex and nonsymmetrical, the eigenvalues and the associated eigenvectors will also, in general, be complex. From (11) and (8) we see that the required transformation matrix $[P(z)]$ also performs the role of a similarity transformation to the matrix $[M(z)]$. If the eigenvalues of the matrix $[M(z)]$ are distinct, there exists a matrix $[P(z)]$ that diagonalizes the matrix $[M(z)]$. The eigenvectors belonging to the corresponding distinct eigenvalues then form an orthonormal basis, *i.e.*,

$$\sum_m P_{mn}P_{ms}^* = \delta_{ns}, \qquad n, s = 1, 2, 3, \cdots, \tag{13}$$

where * denotes the complex conjugate and δ_{ns} the Kronecker delta symbol. When $[P]$ is found, the evaluation of $[P]^{-1}D[P]$ is straightforward. Our task now is to solve the infinite set of coupled differential equations (10) with the appropriate boundary conditions.

The matrix integral of the system of (10) is a square matrix $[B(z)]$, the columns of which are n-linearly-independent solutions of the system. Since each column of the matrix $[B(z)]$ satisfies (10), the matrix integral $[B(z)]$ also satisfies the equation

$$D[B] = [Q][B], \qquad [B(0)] = [I], \tag{14}$$

where $[I]$ is the identity matrix. The formal solution of (10) can now be written in the form

$$\{B\} = [B][P(0)]^{-1}\{C\}. \tag{15}$$

We now seek the solution of system (14)[7] which in view of (11) can be written as

$$D[B] = [q][B] - [\epsilon][B], \qquad [B(0)] = [I]. \tag{16}$$

Under the condition that $[\epsilon]$ may be considered as a very small perturbing matrix relative to the diagonal matrix $[q]$, we can take $[\epsilon][B]$ as the perturbing nonhomogeneous term of the matrix differential equation (16). The approximate solution of (16) in matrix representation is

$$[B] = \exp\left\{\int_0^z [q]dz\right\} - \int_0^z \exp\left\{\int_0^z [q]dz\right\} \cdot \exp\left\{-\int_0^{z'} [q(z')]dz'\right\}[\epsilon(z')][B(z')]dz'. \tag{17}$$

Eq. (17) is a Volterra integral equation of the second kind and can be solved by an iteration method. We obtain, therefore, the final solution of (16) as an infinite series in powers of $[\epsilon]$

$$[B] = \exp\left\{\int_0^z [q]dz\right\}\left\{[I] - \int_0^z \exp\left\{-\int_0^{z'} [q(z')]dz'\right\}[\epsilon(z')]\exp\left\{\int_0^{z'} [q(z')]dz'\right\}dz' + \int_0^z \left[\exp\left\{-\int_0^{z'} [q(z')]dz'\right\}[\epsilon(z')]\exp\left\{\int_0^{z'} [q(z')]dz'\right\}\right] \cdot \left(\int_0^{z'} \exp\left\{-\int_0^{z''} [q(z'')]dz''\right\}[\epsilon(z'')]\exp\left\{\int_0^{z''} [q(z'')]dz''\right\}dz''\right)dz' + \text{terms of higher powers in } [\epsilon]\right\}. \tag{18}$$

[7] See Appendix I.

The uniform convergence of the series in (18) in the closed interval can be shown according to standard methods.[8] The final matrix solution $\{A(z)\}$ is obtained through (18), (15), and (8).

The above general formulation can, in principle, be applied to any number of coupled modes; however, the explicit solutions of the eigenvalues of (12) are already too clumsy to handle even in the case of three coupled modes. Accordingly, only the solution for two coupled modes (TE_{01} and TE_{02}), which have the strongest coupling, will be carried out in the following equation for a gradual taper. Inspection of the coupling coefficient in (5) justifies the preceding statement in addition to the fact that the phase-constant difference is much larger for all other higher-order modes. For TE_{01} and TE_{02} coupling, (12) yields the two eigenvalues

$$\begin{aligned}\lambda_1 &= \tfrac{1}{2}j[-(\beta_1+\beta_2)+\sqrt{(\beta_1-\beta_2)^2+4C_{12}^2}]\\ &= -j(\beta-\Gamma)\end{aligned} \tag{19}$$

and

$$\begin{aligned}\lambda_2 &= \tfrac{1}{2}j[-(\beta_1+\beta_2)-\sqrt{(\beta_1-\beta_2)^2+4C_{12}^2}]\\ &= -j(\beta+\Gamma),\end{aligned}$$

where

$$\Gamma=\sqrt{\Delta\beta^2+C_{12}^2} \tag{20}$$

and

$$\Delta\beta=\tfrac{1}{2}(\beta_1-\beta_2);\qquad \beta=\tfrac{1}{2}(\beta_1+\beta_2).$$

$$[B]=\begin{bmatrix} \exp\left\{-j\int_0^z(\beta-\Gamma)dz\right\} & \\ -j\exp\left\{-j\int_0^z(\beta+\Gamma)dz\right\}\int_0^z\dfrac{d\theta}{dz'} & \\ \cdot\exp\left\{2j\int_0^{z'}\Gamma(z')dz'\right\}dz' & \end{bmatrix}$$

The transformation matrix $[P]$ then takes the form

$$[P]=\begin{bmatrix} j\sqrt{\dfrac{\Gamma+\Delta\beta}{2\Gamma}} & \sqrt{\dfrac{\Gamma-\Delta\beta}{2\Gamma}}\\ \sqrt{\dfrac{\Gamma-\Delta\beta}{2\Gamma}} & j\sqrt{\dfrac{\Gamma+\Delta\beta}{2\Gamma}}\end{bmatrix}. \tag{21}$$

We see that the unitary condition of (13) is satisfied by the $[P]$ matrix, and its determinant is -1 in agreement with the fact that the determinant of an orthonormal matrix can only be $+1$ (corresponding to rotation) or -1 (corresponding to reflection). If use is made of (20) by letting the angle between the sides $\Delta\beta$ and Γ be (2θ), so that $\cos 2\theta=\Delta\beta/\Gamma$, then the matrix $[P]$ of (21) takes the simple form

$$[P]=\begin{bmatrix} j\cos\theta & \sin\theta\\ \sin\theta & j\cos\theta\end{bmatrix}. \tag{22}$$

For small angle (2θ), the variation of (2θ) is directly proportional to that of the coupling coefficient C_{12} and inversely proportional to that of the phase-constant difference $\Delta\beta$ along the taper. Accordingly, this variation of (2θ) can appropriately be interpreted as "mode-conversion distribution" along the taper. The matrix $[\epsilon]$ of (11) in terms of θ becomes a symmetric matrix

$$[\epsilon]=[P]^{-1}D[P]=-j\begin{bmatrix} 0 & \dfrac{d\theta}{dz}\\ \dfrac{d\theta}{dz} & 0\end{bmatrix}. \tag{23}$$

It is clear now that the coupling to the TE_{01} mode becomes much smaller for higher-order modes. For a gradual taper, we see that $[\epsilon]$ is, indeed, a small perturbing term to the diagonal matrix

$$[q]=[P]^{-1}[M][P]=-j\begin{bmatrix}\beta\mp\Gamma & 0\\ 0 & \beta+\Gamma\end{bmatrix}. \tag{24}$$

Note that the trace of the diagonalized matrix remains equal to $(\beta_1+\beta_2)$. For a solution to the first order of approximation, we obtain from (18) via (23) and (24)

$$\begin{bmatrix} -j\exp\left\{-j\int_0^z(\beta-\Gamma)dz\right\}\int_0^z\dfrac{d\theta}{dz'} & \\ \cdot\exp\left\{-2j\int_0^{z'}\Gamma(z')dz'\right\}dz' & \\ \exp\left\{-j\int_0^z(\beta+\Gamma)dz\right\} & \end{bmatrix}. \tag{25}$$

The normalized boundary condition for the two-mode case requires that

$$\begin{aligned}|\{A(0)\}| &= |\{B(0)\}| = |[P(0)]^{-1}\{C\}|\\ &=\begin{bmatrix}1\\0\end{bmatrix},\end{aligned} \tag{26}$$

where the vertical bars denote the norm of a vector of complex elements. For a gentle, smooth taper with a mode-conversion distribution function vanishing at both ends of the taper, we obtain the explicit solution for $A_2(l)$ from (8), (22), (15), (25), and (26)

$$\begin{aligned}A_2(l) &= \exp\left\{-j\int_0^l(\beta+\Gamma)dz\right\}\int_0^l\frac{d\theta}{dz}\\ &\quad\cdot\exp\left\{2j\int_0^z\Gamma(z')dz'\right\}dz.\end{aligned} \tag{27}$$

[8] F. G. Tricomi, "Integral Equations," Interscience Publishers, Inc., New York, N. Y., p. 10; 1957.

Note that $|A_2(l)|$ is essentially in the form of a Fourier transform of $d\theta/dz$. Letting

$$\rho(z) = \int_0^z \Gamma(z')dz' \tag{28}$$

and integrating (27) by parts, we have

$$|A_2(\rho_1)| = \left| \int_0^{\rho_1} (2\theta)e^{2j\rho}d\rho \right|. \tag{29}$$

If we integrate (27) by parts in another way, the following expression is obtained:

$$\begin{aligned} |A_2(\rho_1)| = \Bigg| \frac{1}{2} \Bigg\{ & \left[e^{2j\rho_1} \left(\frac{d\theta}{d\rho} \right)_{\rho=\rho_1} - \left(\frac{d\theta}{d\rho} \right)_{\rho=0} \right] \\ & - \frac{1}{2j} \left[e^{2j\rho_1} \left(\frac{d^2\theta}{d\rho^2} \right)_{\rho=\rho_1} - \left(\frac{d^2\theta}{d\rho^2} \right)_{\rho=0} \right] \\ & + (-1)^{n+1} \left(\frac{1}{2j} \right)^n \left[e^{2j\rho_1} \left(\frac{d^n\theta}{d\rho^n} \right)_{\rho=\rho_1} \right. \\ & \left. - \left(\frac{d^n\theta}{d\rho^n} \right)_{\rho=0} \right] + \cdots \Bigg\} \Bigg|, \end{aligned} \tag{30}$$

where

$$\rho_1 = \rho(l) = \int_0^l \Gamma(z)dz. \tag{31}$$

Eq. (29) is in a form suitable for the computation of "mode discrimination" when the "conversion-distribution function" (2θ) is given in terms of the parameter ρ. Eq. (30), on the other hand, gives us a clue that in order to obtain a higher-mode discrimination, it is advisable to choose a conversion-distribution function with vanishing first and higher derivatives at both taper ends. A detailed discussion will be given in Section V in this respect. When the distribution function (2θ) is chosen, the waveguide radius $a(\rho)$ can be obtained from (5) and (20) and the taper length of the guide $z(\rho)$ from (28).

III. The Choice of Conversion Distribution Function

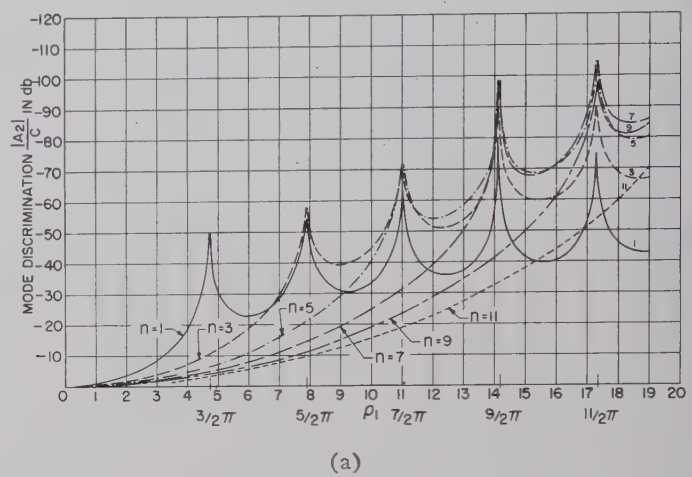
It is evident that the choice of distribution function is not unique. Under the stipulation that the function itself vanishes at both ends as expressed in (8), (22) and (26), a simple choice of such a function in the form of an infinite series in (ρ/ρ_1) is

$$2\theta = K_n \sin^n \left(\frac{\pi\rho}{\rho_1} \right), \tag{32}$$

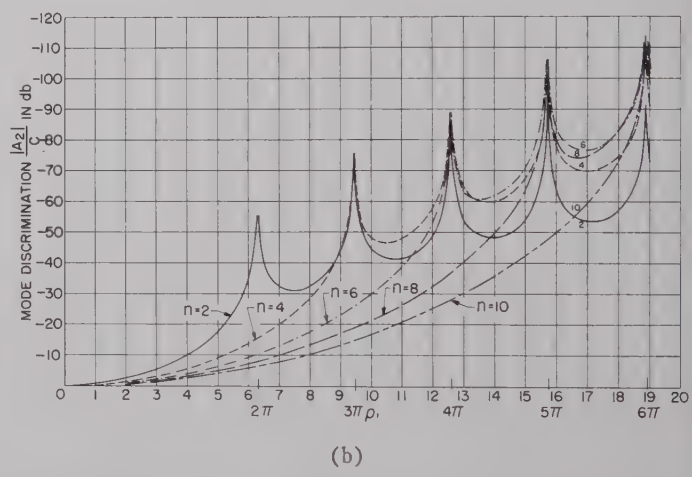
where $\sin(\pi\rho/\rho_1)$ can be considered as a "generating" function, and n may or may not be an integer. Substituting (32) into (29) and using (47) in Appendix II, we can show that the mode discrimination is given by

$$|A_2(\rho_1)| = \left| C(2\cdot 4\cdot 6\cdot \cdots n)^2 \frac{ab^n(e^{a\rho_1} - 1)}{\rho_1 \prod_{s=0}^{n} (a^2 + s^2b^2)} \right|, \quad n \text{ even} \tag{33a}$$

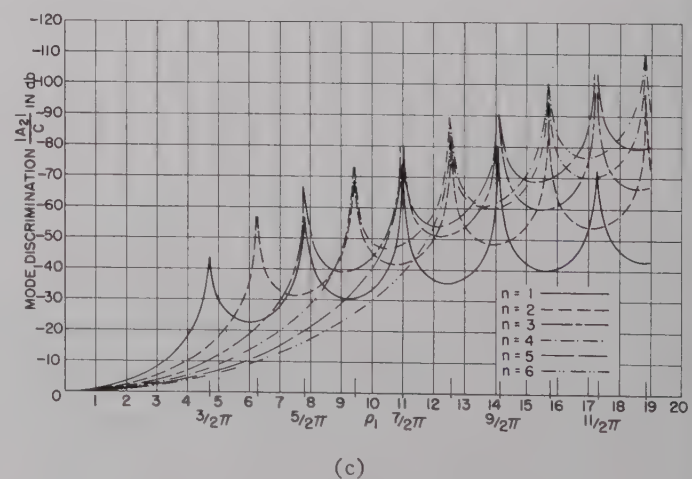
$$|A_2(\rho_1)| = \left| C(1\cdot 3\cdot 5\cdot \cdots n)^2 \frac{b^{n+1}(e^{a\rho_1} + 1)}{2 \prod_{s=1}^{n} (a^2 + s^2b^2)} \right|, \quad n \text{ odd}, \tag{33b}$$



(a)



(b)



(c)

Fig. 2—(a)–(c) Mode conversion in waveguide tapers.

where

$$a = 2j, \qquad b = \frac{\pi}{\rho_1}, \quad \text{and} \quad C = \frac{2k_m k_n}{k_n^2 - k_m^2} \log \frac{a_2}{a_1}, \tag{34}$$

and k_m and k_n are defined in (2). It is evident from (33) that C corresponds to the mode conversion of a step discontinuity in the diameter of the waveguide. For cases where n is noninteger, general solutions for the mode discrimination in closed form are impossible, and we have to resort to numerical integration. Eq. (33) for integral values of n are plotted in Fig. 2(a)–2(c).

It is seen from these figures that there is always an "optimum" integer n which minimizes the length of a taper for a prescribed discrimination. Alternatively, if we prescribe the taper length, there is always an optimum integer n that provides the highest-mode discrimination. From Fig. 2(a) we see that, for a fixed taper length of $\rho_1 = 18.5$, optimum discrimination occurs for $n = 7$. On the other hand, if we prescribe a -50-db mode discrimination within a frequency range up to 75 kMc, the required minimum length of the taper for the case ($a(0) = 7''/16$ and $a(l) = 1''$) is about 3 ft[4] using $n = 2$, but only about 2 ft using $n = 3$ as shown in Fig. 3. The actual computation of radius a and length z is shown in Appendix II.

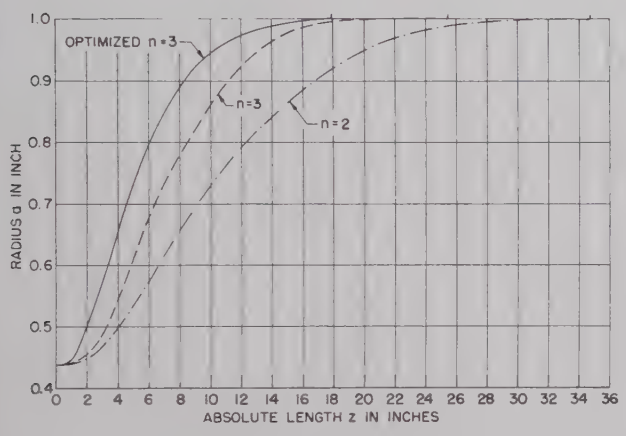


Fig. 3—Comparison of profile of tapers of same mode conversion (50 db).

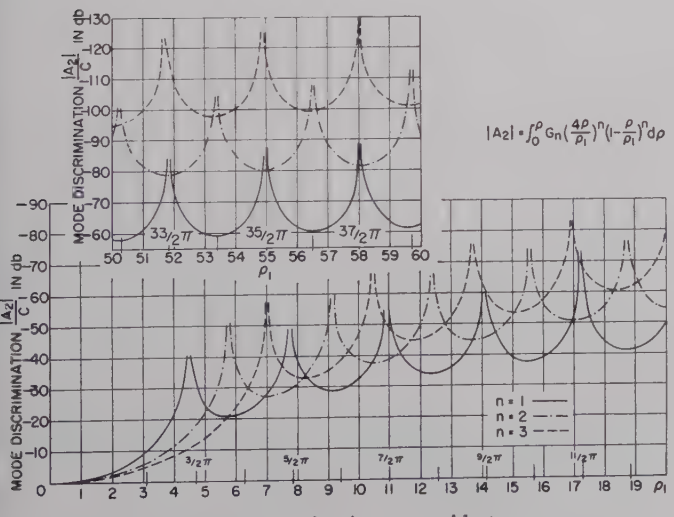


Fig. 4—Mode conversion in waveguide tapers.

Another simple choice for the distribution function is a polynomial in (ρ/ρ_1)

$$2\theta = G_n \left(\frac{4\rho}{\rho_1}\right)^n \left(1 - \frac{\rho}{\rho_1}\right)^n, \tag{35}$$

where the "generating" function is $(4\rho/\rho_1)(1-\rho/\rho_1)$, and n may or may not be an integer. Evaluating the integral of (29) for (35) with integer n, it can be shown that the mode discrimination is equal to

$$\begin{aligned}
|A_2(\rho_1)| &= \left| \frac{3 \cdot C}{2 \cdot \rho_1} \left(\frac{2}{\rho_1}\right)^2 \left[\left(\frac{\rho_1}{a^2}(e^{a\rho_1} + 1) - \frac{2}{a^3}(e^{a\rho_1} - 1)\right)\right] \right| \quad \text{for } n = 1, \\
|A_2(\rho_1)| &= \left| \frac{3 \cdot 5 \cdot C}{2^3 \cdot \rho_1} \left(\frac{2}{\rho_1}\right)^4 \left[\left(\frac{2\rho_1^2}{a^3} + \frac{24}{a^5}\right)(e^{a\rho_1} - 1) - \frac{12\rho_1}{a^4} \cdot (e^{a\rho_1} + 1)\right] \right| \quad \text{for } n = 2, \\
|A_2(\rho_1)| &= \left| \frac{5 \cdot 7 \cdot C}{2^4 \cdot \rho_1} \left(\frac{2}{\rho_1}\right)^6 \left[\frac{6\rho_1}{a^4}\left(\rho_1^2 + \frac{60}{a^2}\right)(e^{a\rho_1} + 1) - \frac{72}{a^5} \cdot \left(\rho_1^2 + \frac{10}{a^2}\right)(e^{a\rho_1} - 1)\right] \right| \quad \text{for } n = 3,
\end{aligned} \tag{36}$$

where C is defined in (34). Eq. (36) is plotted in Fig. 4. It can be shown that both (33) and (36) have the value unity times C at $\rho_1 = 0$, as they should. Eqs. (33) have all their zeros at multiples of $\pi/2$; on the other hand, (36) has its initial zeros shifted closer toward the origin $\rho_1 = 0$, as shown in Fig. 4. Inspection of Fig. 4 shows that the zeros at large values of ρ_1 also gradually shift to positions at multiples of $\pi/2$ as (33) do. Comparison of Fig. 2 and Fig. 4 for curves of corresponding values of n shows that the sine distribution has better over-all discrimination than the polynomial distribution, except in the region between $\rho_1 = 0$ and the first zero of (33).

IV. Optimization

In an attempt to further generalize and optimize the mode-conversion distribution function, we expand it in a symmetrical Fourier series (with the origin of the coordinate system shifted to the center of the taper)

$$2\theta = \sum_{m=0}^{\infty} D_m \cos m\left(\frac{\pi\rho}{\rho_1}\right). \tag{37}$$

With suitably-chosen coefficients D_m's, it is obvious that

(37) can include all cases described by (32) and (35). Substituting (37) into (29), we have

$$|A_2(\rho_1)| = \left| \int_{-\rho_1/2}^{\rho_1/2} \sum_{m=0}^{\infty} D_m \cos m\left(\frac{\pi\rho}{\rho_1}\right) e^{2j\rho} d\rho \right|$$

$$|A_2(\rho_1)| = \left| \sum_{m=0}^{\infty} \frac{2D_m \cos m\frac{\pi}{2}(e^{a\rho_1}-1)}{a^2+m^2b^2} \right|, \quad m \text{ even} \quad (38a)$$

$$|A_2(\rho_1)| = \left| \sum_{m=1}^{\infty} \frac{mbD_m \sin m\frac{\pi}{2}(e^{a\rho_1}+1)}{a^2+m^2b^2} \right|, \quad m \text{ odd} \quad (38b)$$

where a and b are defined in (34). It is important to note that only odd or even values of m are required to represent the mode-conversion distribution function given in (33), depending on whether n is odd or even. Equating (38) to (33) for the corresponding case, we can obtain the D_m in (38) as shown in Table I. It is interesting to note that the ratio of D_m's for a particular n correspond to the coefficients of the binomial expansion. We are now in a position to further optimize the mode-conversion curve defined by (33), using the values of Table I as a guide. The aim is to reshape the mode-conversion curve in such a way that the first few maxima of the "side lobes" of (33) will be leveled and at the same time lowered optimally. Before carrying out the optimizing procedure, it is appropriate at this point to discuss the incentive for this further optimization in more detail in the following paragraph.

TABLE I

$n=0$	$D_0=\frac{C}{\rho_1}$.		
$n=2$	$D_0=\frac{C}{\rho_1}$,	$D_2=\frac{C}{\rho_1}$.	
$n=4$	$D_0=\frac{C}{\rho_1}$,	$D_2=\frac{4}{3}\cdot\frac{C}{\rho_1}$,	$D_4=\frac{1}{4}\cdot\frac{C}{\rho_1}$.
⋮	⋮	⋮	⋮
$n=1$	$D_1=\frac{\pi C}{2\rho_1}$.		
$n=3$	$D_1=\frac{9}{8}\cdot\frac{\pi C}{2\rho_1}$,	$D_3=\frac{1}{3}\cdot\frac{9}{8}\cdot\frac{\pi C}{2\rho_1}$.	
$n=5$	$D_1=\frac{75}{64}\cdot\frac{\pi C}{2\rho_1}$,	$D_3=\frac{1}{2}\cdot\frac{75}{64}\cdot\frac{\pi C}{2\rho_1}$,	$D_5=\frac{1}{10}\cdot\frac{75}{64}\cdot\frac{\pi C}{2\rho_1}$.
⋮	⋮	⋮	⋮

The class of distribution functions that we have been considering have the property that the distribution function and all its derivatives are single-valued, uniformly bounded, and continuous in the interval of interest. Any distribution function of this class can be transformed to a general form in terms of the zeros of the function. With the coordinate origin at the center of the taper, the function of interest has the form

$$2\theta = f(\rho)\left[\rho^2 - \left(\frac{\rho_1}{2}\right)^2\right]^n, \quad (39)$$

where $f(\rho)$ is an even function due to the symmetry of the function (2θ), and $f(\rho)$ does not vanish at the taper ends $(\rho=\pm\rho_1/2)$ or any other value of ρ in the interval. Taylor[9] has shown that the Fourier transform of (39) has the following asymptotic form as ρ_1 approaches infinity:

$$F(\rho_1) \sim f\left(\frac{\rho_1}{2}\right)\Gamma(n+1)\frac{\cos\left(\rho_1-\frac{n+1}{2}\pi\right)}{(\rho_1)^{n+1}}. \quad (40)$$

It is, therefore, seen that the mode discrimination of very long tapers is only trivially different, no matter what form the distribution function has. On the other hand, as we have seen earlier, the initial slope of the function and accordingly the value of n are of considerable importance in "optimizing" the taper. The initial slope of the function (2θ) determines the positions of the zeros near the origin of the mode discrimination curve $A_2(\rho)$. The smaller the initial slope, the further will the first zero be from the origin. Likewise, the value of n alters the zero positions, since n changes the slope of the function. Eq. (40) indicates clearly that spurious mode conversion or discrimination decays as $1/|\rho_1|^{n+1}$ and zeros appear at $\rho_1=n(\pi/2)$ at large values of ρ_1. Inspection of (33), (36) and (38) confirms this decay rate and the position of zeros at large values of ρ_1. At this point, it is particularly appropriate to compare this decay rate with the nondecaying characteristics of the Tchebycheff polynomial of infinite degree. If we are to simulate the Tchebycheff polynomial of infinite degree by the function of (40), we see that the only choice to make (40) nondecaying is to make $n=-1$. However, for this choice of n, the function of (39) will have poles at the taper ends and will no longer be uniformly bounded. This violates our basic requirement, and, therefore, it is clear that a smooth transition taper with its mode-conversion characteristics described by a Tchebycheff polynomial of infinite degree is unrealizable. In fact, this unrealizability is simply a consequence of the law of conservation of energy. This is why even

[9] T. T. Taylor, "Design of line-source antenna for narrow bandwidth and low side lobes," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-3, pp. 16–28; January, 1955.

in the transmission line case either steps[10] have to be introduced at taper ends or a modified[11] Tchebycheff polynomial of infinite degree has to be used in the synthesis procedure. Although the frequency range of the design using infinite degree Tchebycheff polynomial extends to infinity, for band-pass applications this is unnecessary. Accordingly, the Tchebycheff design is optimum only in the sense of infinite bandwidth, even for a transmission-line taper.

From the above exposition, we see that in order to further optimize, we need to "flatten" or "level" the decay rate of the first few "minor lobes" of the mode-conversion curve as much as permissible, after choosing the "optimum" value of n for a prescribed discrimination level from Fig. 2(c). It is clear that the nature of this procedure essentially becomes that of synthesis. At first sight, this might look rather aimless if we do it in a heuristic way by adjusting the coefficients D_m of (38). Knowing, however, that the shape of the mode-conversion curve of (38) depends very much on the density of the distribution of zeros near the origin, we are thus led to create an extra "zero" at such a position that the "side lobes" near the origin will be leveled and at the same time optimally lowered. It is important to note that the new zeros will be introduced by properly choosing the coefficients D_m in (38), while the original zeros were determined only by the term $(e^{a\rho_1} \pm 1)$. This will be done first for $n=3$. With reference to (38b), we have

$$D_1'\left[3^2 - \left(\rho_1 \frac{2}{\pi}\right)^2\right] - 3D_3'\left[1^2 - \left(\rho_1 \frac{2}{\pi}\right)^2\right] = 0, \qquad (41)$$

where

$$D_m = \frac{bC}{2} D_m' = \frac{\pi C}{2\rho_1} D_m'.$$

With (41) and the boundary condition that $A_2(0) = C$,

$$\left(D_1' - \frac{D_3'}{3}\right) = 1. \qquad (42)$$

We can solve for the required coefficients D_1' and D_3' when the position of the new zero is intelligently selected. Inspection of Fig. 2(a) for $n=3$ shows that it is advisable to locate the "new zero" around $\rho_1 = 6(\pi/2)$ in order to achieve the desired results. With this value of ρ_1, we get $D_1' = 1.09375$ and $D_3' = 0.28125$ from (41) and (42). A plot of (38) with these values of D_1' and D_3' as shown in Fig. 5 indeed gives the desired results. Fig. 6, showing the relation between the position of the new zero and the maximum of the maxima of the side lobes, also confirms the fact that the optimum location of the new zero should be in the vicinity of $6(\pi/2)$. Mode-conversion curves for

$$\cos^3\left(\frac{\pi\rho}{\rho_1}\right) \quad \text{and} \quad \left[\frac{\rho}{\rho_1}\left(1 - \frac{\rho}{\rho_1}\right)\right]^3$$

are also plotted in Fig. 5 for comparison. It is seen that the optimized conversion curve does have its first few side lobes maxima leveled, and the maxima of the remaining side lobes decay according to a rate asymptotically proportional to $1/\rho_1^4$. The improvement is about another 30 per cent reduction in taper length for 50-db discrimination. The actual optimized length of the taper is plotted in Fig. 3 for comparison. Through using the optimized $\cos^3(\pi\rho/\rho_1)$ distribution function, a total reduction of about 50 per cent in taper length is realized for the case of 50-db prescribed mode discrimination in a taper connecting a $\frac{7}{8}$-in ID waveguide to a 2-in ID waveguide.

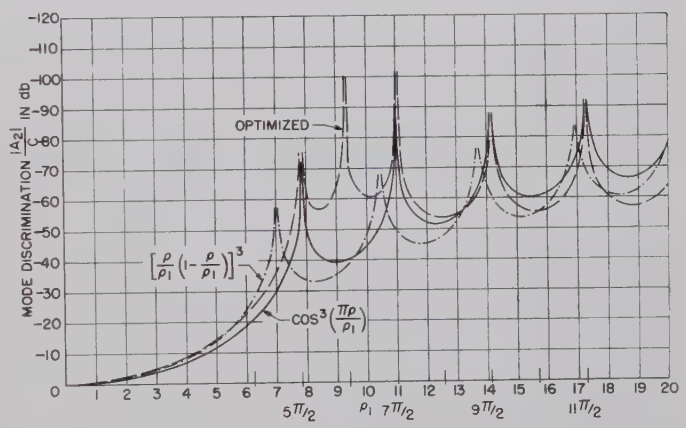


Fig. 5—Comparison of mode conversion in waveguide tapers.

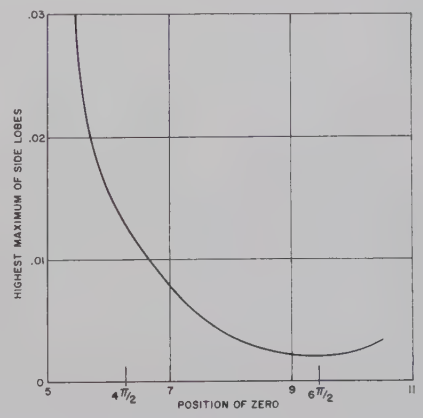


Fig. 6—Optimization of the position of new zero.

[10] R. W. Klopfenstein, "A transmission line taper of improved design," PROC. IRE, vol. 44, pp. 31–35; January, 1956.
[11] R. E. Collin, "The optimum tapered transmission line matching section," PROC. IRE, vol. 44, pp. 539–548; April, 1956.

The same procedure can be applied, for example, to the $n=5$ case and we have

$$D_1'\left[3^2-\left(\rho_1\frac{2}{\pi}\right)^2\right]\left[5^2-\left(\rho_1\frac{2}{\pi}\right)^2\right]$$

$$-3D_3'\left[1^2-\left(\rho_1\frac{2}{\pi}\right)^2\right]+\left[5^2-\left(\rho_1\frac{2}{\pi}\right)^2\right]$$

$$+5D_5'\left[1^2-\left(\rho_1\frac{2}{\pi}\right)^2\right]\cdot\left[3^2-\left(\rho_1\frac{2}{\pi}\right)^2\right]=0 \qquad (43)$$

and

$$\left[D_1'-\frac{D_3'}{3}+\frac{D_5'}{5}\right]=1, \qquad (44)$$

where

$$D_m=\frac{\pi C}{2\rho_1}D_m'.$$

Inspection of Fig. 2(a) shows that the logical choice for the location of the new zero is at $\rho_1=8(\pi/2)$. In this case, however, we have only two equations for three unknowns. It is necessary to assume an appropriate value of D_1' so that D_3' and D_5' can be determined. The first suitable choice of D_1' might be $D_1'=75/64$ obtained from Table I for the $n=5$ case. Calculation shows again that the new zero should indeed be around $\rho_1=8(\pi/2)$, and the optimum choice of D_1 is indeed (75/64). This case is shown in Fig. 7 together with the $\cos^5(\pi\rho/\rho_1)$ case for comparison.

Further investigation of the $n=5$ case reveals a better value for D_1' because we can now create two new zeros in view of the extra undetermined coefficient. The assignment of the second zero will give us an extra equation and thus determine the three coefficients uniquely. A judicial choice of two zeros at $\rho_1=8(\pi/2)$ and $\rho_1=10(\pi/2)$, respectively, should yield even better mode discrimination at a fixed taper length than the above case. This result is also plotted in Fig. 7 for comparison and shows that the prediction is valid. A relocation of the first "created zero" at $\rho_1=7.72(\pi/2)$ gives the best discrimination for this case, as shown in Fig. 7. Accordingly, it is seen that we can further optimize a taper by creating new zeros near the first few zeros of the mode-conversion curve, and the number of new zeros allowed to be created increases directly as the number of undetermined coefficients. On the other hand, the freedom to have more created zeros is only available for higher values of n which dictates higher mode discrimination and longer taper lengths.

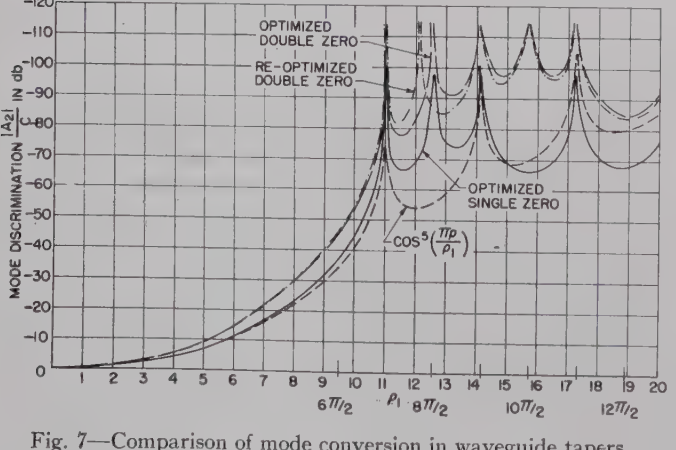


Fig. 7—Comparison of mode conversion in waveguide tapers.

V. Conclusions

In an attempt to approximate a Tchebycheff mode-conversion response in a wideband waveguide taper, we optimize the taper by creating new zeros in the mode-conversion response. This response had been initially selected to yield the "shortest" taper length in the "sine or cosine distribution function family" at a prescribed level of mode discrimination.

We note that the first and second derivative of a distribution function with ratio $D_3/D_1=\frac{1}{3}$ [this is the $\cos^3(\pi\rho/\rho_1)$ case] vanish at the taper ends, but those with ratios other than $\frac{1}{3}$ do not vanish. Thus, change of end slopes evidently is a consequence of the addition of new zeros in the mode-conversion curve, since the end slopes are closely related to the position and density of the zero distribution near the origin. Because the optimized distribution functions do not have vanishing derivatives at the taper ends, it is seen that the distribution functions with vanishing derivatives at the taper ends may not be most desirable. We now return to (30) for further information in this respect. It can be easily shown that for the symmetrical distribution functions we used, the nth derivatives at the taper ends decreases as $1/\rho_1^{n+1}$, where ρ_1 is the equivalent taper length. This implies that longer tapers have smaller derivatives and that the $(n+1)$th derivative is smaller than the nth derivative. Investigation of (30) also shows that if a symmetrical distribution function has its first and second derivative vanishing, the third derivative will represent a significant part of the mode conversion. On the other hand, if the first two derivatives are nonvanishing but very small, the total mode conversion may still be smaller than the vanishing derivative case due to alternate $+$ signs and $-$ signs in the real and imaginary parts of (30).

From what has been shown here, it is obvious that we cannot claim to have synthesized the absolute optimum taper, but we can claim the taper to be very close to the absolute optimum. The procedure discussed above

may lead to a further optimization, but it is quite clear that the small reduction of taper length due to such a procedure might not justify the amount of computational labor involved. The starting point is still (32). For instance, for a certain prescribed spurious mode level, there is in (32) an "optimum" n which may be noninteger. The evaluation of (29) with a distribution function of noninteger power will require numerical integration, and the evaluation of the radius and length of the taper as shown in Appendix II will again require numerical integration.

Appendix I

If the matrix $[\delta]$ can be considered as a perturbing matrix to a matrix $[N]$, the matrix exponential $e^{[N]+[\delta]}$ can be expressed in several ways. If $[N]$ and $[\delta]$ commute, then

$$e^{[N]+[\delta]} = e^{[N]}e^{[\delta]}.$$

Considering the general case where $[N]$ and $[\delta]$ do not commute, we can write

$$e^{[N]+[\delta]} = [I] + \sum_{k=1}^{\infty} \frac{([N] + [\delta])^k}{k!}.$$

By expanding the series, we can collect the terms in powers of the perturbing term $[\delta]$ in the form of

$$e^{[N]+[\delta]} = e^{[N]} + \sum_{k=1}^{\infty} [\delta]^k f_k([N]).$$

This, however, cannot be accomplished in a neat fashion, and it is necessary to solve equations of the form of (16) by perturbation techniques.

Appendix II

The coefficient C_{12} is obtained from (5) under the stipulation of the TE_{01} and TE_{02} modes that are far away from cut-off, and it takes the following simple form:

$$C_{12} = \frac{1}{a}\frac{da}{dz}\cdot\frac{2k_1k_2}{k_2^2 - k_1^2} = \frac{k}{a}\frac{da}{dz},$$

where

$$k = \frac{2k_1k_2}{k_2^2 - k_1^2}.$$

From the preceding equation and (20) and (28), we have

$$\int_{a_1}^{a} \frac{da}{a} = \int_0^{\rho} \frac{2\theta}{k}\, d\rho \qquad (45)$$

for a gentle taper.

For $2\theta = K_n \sin^n(\pi\rho/\rho_1)$, (45) becomes

$$\log\frac{a}{a_1} = \frac{K_n}{k}\left[-\frac{\sin^{n-1}\left(\frac{\pi\rho}{\rho_1}\right)\cos\left(\frac{\pi\rho}{\rho_1}\right)}{n\left(\frac{\pi}{\rho_1}\right)} + \frac{n-1}{n}\int_0^{\rho}\sin^{n-2}\left(\frac{\pi\rho}{\rho_1}\right)d\rho\right]. \qquad (46)$$

Boundary conditions require that

$$K_n = \frac{1}{2}\left[1\cdot\frac{3}{2}\cdot\frac{5}{4}\cdots\frac{n}{n-1}\frac{\pi k\log\frac{a_2}{a_1}}{\rho_1}\right], \quad n \text{ odd}$$

$$K_n = \frac{2}{1}\cdot\frac{4}{3}\cdot\frac{6}{5}\cdots\frac{n}{n-1}\frac{k\log\frac{a_2}{a_1}}{\rho_1}, \quad n \text{ even}. \qquad (47)$$

Substitution of (47) into (46) yields for

$$\frac{1}{2}\left[1 - \cos\left(\frac{\pi\rho}{\rho_1}\right)\right]\log\frac{a_2}{a_1}$$

$n = 1$: $a = a_1 e$

$$\left[\frac{\rho}{\rho_1} - \frac{1}{2\pi}\sin 2\pi\frac{\rho}{\rho_1}\right]\log\frac{a_2}{a_1}$$

$n = 2$: $a = a_1 e$

$$\frac{1}{2}\left[1 - \frac{3}{2}\cos\left(\frac{\pi\rho}{\rho_1}\right) + \frac{1}{2}\cos^3\left(\frac{\pi\rho}{\rho_1}\right)\right]\log\frac{a_2}{a_1}$$

$n = 3$: $a = a_1 e$. (48)

The actual length of the taper in terms of the parameters ρ_1 is obtained from (28). For a gentle taper it is

$$z = \frac{1}{k_2^2 - k_1^2}\left[4\beta_0\int_0^{\rho} a^2 d\rho - \frac{\rho}{\beta_0}(k_2^2 + k_1^2) - \frac{1}{4\beta_0^3}(k_2^4 + k_1^4)\int_0^{\rho}\frac{d\rho}{a^2}\right], \qquad (49)$$

where β_0 is the phase constant in free space. The last term in (49) is negligibly small, in general. Substitution of (48) into (49) gives

$$n = 1: \quad z = \frac{4\beta_0 a_1/a_2\rho_1}{\pi(k_2^2 - k_1^2)}\left[\left(1 + \frac{\alpha^2}{4} + \frac{\alpha^4}{64}\right)x - \left(\alpha + \frac{\alpha^5}{150}\right)\sin x + \left(\frac{\alpha^2}{8} + \frac{\alpha^4}{128}\right)\sin 2x + \left(\frac{\alpha^3}{18} + \frac{\alpha^5}{450}\right)\sin^3 x + \frac{\alpha^4\cos^3 x\sin x}{96} - \frac{\alpha^5\cos^4 x\sin x}{600}\right] - \frac{\rho(k_2^2 - k_1^2)}{\beta_0},$$

where

$$x = \frac{\pi\rho}{\rho_1} \quad \text{and} \quad \alpha = \log_e \frac{\alpha_2}{\alpha_1}$$

$n = 2$:

$$z = \frac{2a_1^2\beta_0\rho_1}{\pi(k_2^2 - k_1^2)} \left\{ e^{qy} \left[\frac{1}{q} - \left(\frac{8q + q^3}{8q^2 + 8} \right) (q \sin y - \cos y) \right. \right.$$

$$+ \frac{12q + q^3}{48} - \left(\frac{12q^2 + q^4}{q^2 + 4} \right) (q \cos 2y + 2 \sin 2y)$$

$$\left. - \frac{q^4(2 \sin 2y + q \cos 2y + 4 \sin 4y + q \cos 4y)}{192(q^2 + 16)} \right]$$

$$- \left[\frac{1}{q} + \frac{8q + q^3}{8(q^2 + 1)} + \frac{12q + q^3}{48} - \frac{12q^3 + q^5}{q^2 + 4} \right.$$

$$\left. \left. - \frac{q^5}{96} \right] \right\} - \frac{\rho(k_2^2 - k_1^2)}{\beta_0},$$

where

$$y = 2\pi \frac{\rho}{\rho_1} \quad \text{and} \quad q = \frac{1}{\pi} \log_e \frac{a_2}{a_1}.$$

$n = 3$:

$$z = \frac{4\beta_0 a_1 a_2 \rho_1}{\pi(k_2^2 - k_1^2)} \left[x - \frac{3}{2} \alpha \sin x + \frac{9}{16} \alpha^2 \left(\frac{x}{3} + \frac{\sin 2x}{2} \right) \right.$$

$$+ \left(\frac{1}{2} \alpha - \frac{27}{112} \alpha^3 + \frac{81}{1120} \alpha^5 \right) \left(\sin x - \frac{\sin^3 x}{3} \right)$$

$$- \left(\frac{31}{48} \alpha^2 - \frac{99}{256} \alpha^4 - \frac{81}{6144} \alpha^6 \right) \left(\frac{3x}{8} + \frac{3 \sin 2x}{16} \right.$$

$$\left. + \frac{\cos^3 x \sin x}{4} \right) + \left(\frac{\alpha^2}{8} - \frac{81}{512} \alpha^4 + \frac{81}{5120} \alpha^6 \right)$$

$$\cdot \frac{\cos^5 x \sin x}{6} + \frac{9\alpha^4}{64} \frac{\sin x \cos^7 x}{8} + \left(\frac{45}{112} \alpha^3 - \frac{243}{8960} \alpha^5 \right)$$

$$\left. \cdot \frac{\cos^4 x \sin x}{5} + \left(\frac{27}{256} \alpha^5 - \frac{3}{16} \alpha^3 \right) \frac{\sin x \cos 6x}{7} \right] - \frac{\rho(k_2^2 - k_1^2)}{\beta_0}.$$

Acknowledgment

The author is indebted to Miss E. G. Cheatham for her assistance in carrying out most of the computations and to Mrs. M. O. Czarnomski for carrying out the initial phase of computation.

Correspondence

A Broad-Band Coaxial Ferrite Switch*

A broad-band strip-line reflective ferrite switch has been described by Johnson and Wiltse,[1] who also referred to the possibility of a similar switch in coaxial line. This note describes a coaxial on-off switch which will operate over the band 2500–4100 Mc; two such units can be combined to make a two-way switch. An isolation of 40 db was achieved, with a very low loss for the transmitting path, which, in both these devices, was obtained by magnetizing the ferrite well beyond the value for isolation.[2]

Each switching element employed a small slug of a developmental ferrite (B_{sat}=2280 gauss) which completely filled a half-inch section of air spaced coaxial line of 9/32 inch outer diameter and $\frac{1}{8}$ inch inner diameter, the ends of which were directly coupled to Type "C" coaxial connectors. Fig. 1 shows the attenuation obtained with a field of about 400 oe compared with pads of approximately 30 db and 40 db, while Fig. 2 shows the attenuation obtained with a field of about 2500 oe compared with a 3-db pad. The attenuation of about 40 db was obtained at 100°C in a convection cooled solenoid, but greater attenuation was achieved at lower temperatures. The VSWR under reflecting conditions in Fig. 3 is about 0.15, but a lower value may be obtained over a smaller bandwidth. The solenoid power required was rather high, but was reduced by constructing the coaxial line of iron with a brass section in the position of the ferrite slug. Permanent magnet bias can also be used without unduly slowing the switching speed.

Fig. 4 shows the variation in the position of the effective short-circuit planes in front of a slug of a similar ferrite material. A two-way switch may be constructed therefore by arranging for a high impedance to appear at the T junction. Two ferrite slugs were used in the top of the "T," one of which was magnetized for isolation and the other for transmission at any instant. The transmission loss for the two-way switch is shown in Fig. 5.

The development of this switch was part of the work done under a contract for the Admiralty.

* Received by the PGMTT, June 5, 1961.

[1] C. M. Johnson and J. C. Wiltse, "A broad-band ferrite reflective switch," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-8, pp. 466–467; July, 1960.

[2] The General Electric Company, Ltd., Brit. Patent Application No. 19948/59; June 10, 1959.

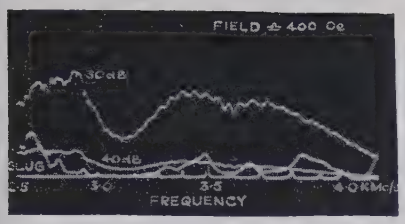


Fig. 1—Isolation of ferrite slug.

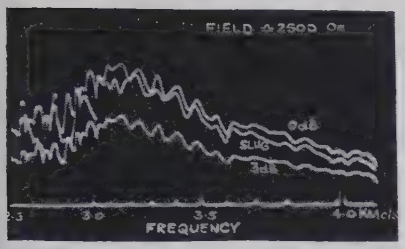


Fig. 2—Transmission loss of ferrite slug.

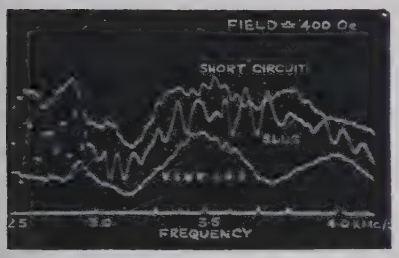


Fig. 3—Reflection coefficient of isolating slug.

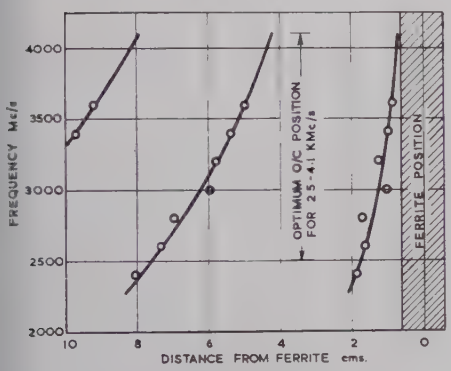


Fig. 4—Short-circuit planes produced under constant reflecting conditions.

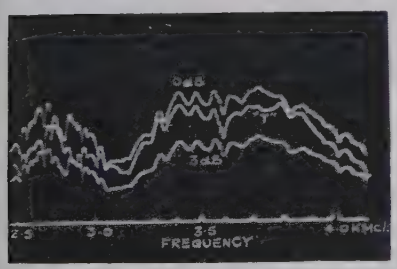


Fig. 5—Transmission loss of T junction.

R. L. Booth
The General Electric Company, Ltd.
Central Research Labs.
Hirst Research Centre
Wembley, England

Conditions for Maximum Power Transfer*

It is sometimes of interest to ask for the conditions of maximum power transfer from a fixed source into a load constrained to vary over an arbitrary contour in the impedance plane. There exists a simple graphical solution to this question as shown below. Consider the circuit shown in Fig. 1.

If P is the power delivered to the load and $P_0=(E_s^2/4R_s)$ is the available power from the source, then

$$\frac{P}{P_0}=\frac{4r}{(r+1)^2+(x+x_s)^2}, \qquad (1)$$

giving

$$\left[r-\left(2\frac{P_0}{P}-1\right)\right]^2+(x+x_s)^2 = \left(2\frac{P_0}{P}-1\right)^2-1, \qquad (2)$$

where

$$r=\frac{R}{R_s}, \quad x=\frac{X}{R_s}, \quad x_s=\frac{X_s}{R_s}, \quad z=\frac{Z}{R_s}, \quad z_s=\frac{Z_s}{R_s}.$$

Eq. (2) represents a family of circles in the z plane of radius $\sqrt{(2P_0/P-1)^2-1}$, whose centers lie along the line $x=-x_s$ through the point z_s^* as shown in Fig. 2.

If a is the distance from unity to the center of a circle along the line $x=-x_s$, the equation for the family of circles becomes

$$[r-(a+1)]^2+(x+x_s)^2=(a+1)^2-1,$$

where $a=2(P_0/P-1)$.

To find the impedance for maximum power transfer to Z, on the plane in which z is drawn, strike a line parallel to the r axis through the point $z_s^*=1-jx_s$. Move an arbitrary distance a from unity along this line, and from this point draw a circle of radius

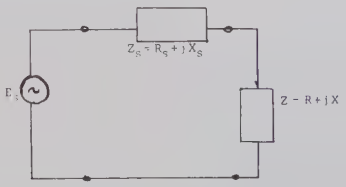


Fig. 1.

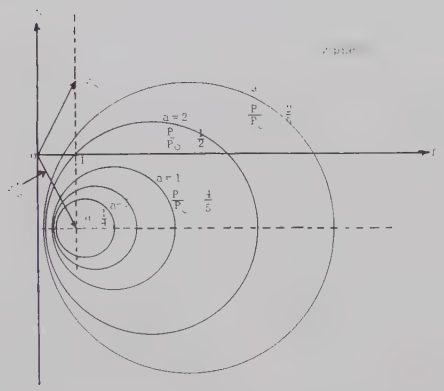


Fig. 2.

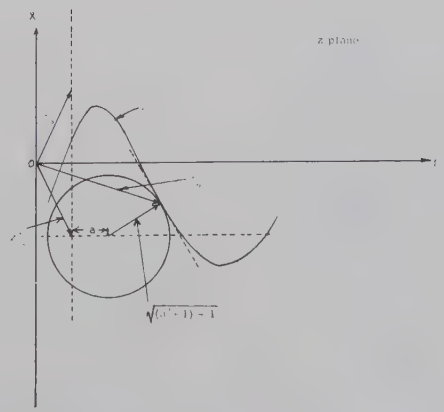


Fig. 3.

* Received by the PGMTT, June 15, 1961.

$\sqrt{(a+1)^2-1}$. Repeat until the circle is just tangent to the given z curve. The point of tangency gives the impedance Z_m for maximum transfer, while

$$\frac{P}{P_0}=\frac{2}{a+2}$$

as shown in Fig. 3.

Note that maximum power transfer does not occur at the point of closest approach to Z_s^*.

CARL SHULMAN
School of Technology
Dept. of Elec. Engrg.
The City College
New York, N. Y.

A Comment on the Scattering Matrix of Cascaded 2*n*-Ports*

Epprecht[1] calculated the scattering matrix of two cascaded two-ports. Redheffer[2] does the same for the $2n$-port using nonstandard notation. This note will comment on the physical interpretation of the constituents of the resultant-scattering matrix. To use the notation of Fig. 1, the scattering matrix constituents are

$$S_{11} = S_{11}' + S_{12}'S_{11}''(1 - S_{22}'S_{11}'')^{-1}S_{21}$$
$$S_{12} = S_{12}'(1 - S_{11}'S_{22}')^{-1}S_{12}''$$
$$S_{21} = S_{21}''(1 - S_{22}'S_{11}'')^{-1}S_{21}'$$
$$S_{22} = S_{22}'' + S_{21}''S_{22}'(1 - S_{11}''S_{22}')^{-1}S_{12}''. \quad (1)$$

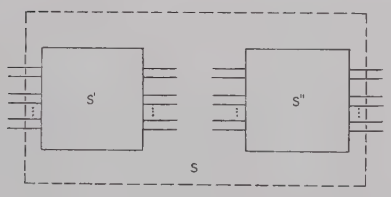


Fig. 1—S' and S'' are $n \times n$ scattering matrices of the respective networks. S is the scattering matrix of the resultant network:

$$S' = \left[\begin{array}{c|c} S_{11}' & S_{12}' \\ \hline S_{21}' & S_{22}' \end{array}\right] \quad S'' = \left[\begin{array}{c|c} S_{11}'' & S_{12}'' \\ \hline S_{21}'' & S_{22}'' \end{array}\right].$$

The interpretation given to these formulas is that S_{11} is the bilinear transformation of S_{11} through the single primed network, and S_{22} is the bilinear transformation of S_{22}'' through the double primed network. Both of these results also follow from the definition of scattering matrix terms on the basis of matched termination (*i.e.*, if the output has a matched load, $S=0$, the input coefficient of the double primed network is S_{11}'', which is the output coefficient of the primed network). S_{12} and S_{21} are similarly interpretable, with the special case of bilaterally matched networks being the "star" multiplication of Altschuler and Kahn.[3]

It should also be noted that formulas (1) are valid when an n-port and an m-port are cascaded[4] (or interconnected).

D. J. R. Stock
L. J. Kaplan
Dept. Elec. Engrg.
New York University
New York, N. Y.

* Received by the PGMTT, June 21, 1961. This research was sponsored by the Electronics Research Directorate of the Air Force Res. Div., Air Res. and Dev. Command, under Contract No. AF19(604)7486.

[1] G. W. Epprecht, "Allgemeine Aktive, Passive und Nichtreziproke Vierpole," *Tech. Mitt. PTT*, NR 5, pp. 169–173; 1959.

[2] R. M. Redheffer, "Inequalities for a matrix Riccati equation:" *J. Math. and Mech.*, vol. 3, pp. 349–367; May, 1959.

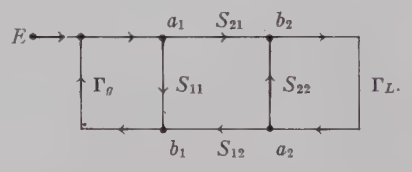
[3] H. M. Altschuler, and W. K. Kahn, "Nonreciprocal two-ports represented by modified Wheeler networks," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-4, pp. 228–233; October, 1956.

[4] L. J. Kaplan. and D. J R. Stock, "A generalization of the matrix Riccati equation and the 'Star' multiplication of Redheffer," *J. Math. and Mech.*, vol. 6; November, 1962.

Use of Flow Graphs to Evaluate Mistermination Errors in Loss and Phase Measurements*

The purpose of this note is to show how the signal flow graph technique illustrated by Hunton[1] leads quite naturally to an expression for error due to mistermination when measuring insertion loss and phase.

We start with the flow graph used by Hunton to represent the tandem connection of generator, network and load:



With the aid of the nontouching loop rule to solve the graph, Hunton easily obtained the result

$$\frac{b_2}{E}=\frac{S_{21}}{1-\Gamma_g S_{11}-\Gamma_L S_{22}+\Gamma_g\Gamma_L(S_{11}S_{22}-S_{12}S_{21})}.$$

Very little extra work is needed to compute insertion loss and phase measurement errors due to mistermination, once the above equation is availabie.

E is the wave amplitude at the output port of the generator when terminated in a matched load Z_0. If V_g znd Z_g represent the Thevenin generator voltage and impedance, then

$$E = \frac{Z_0}{Z_0 + Z_g} V_g.$$

Since

$$Z_g = Z_0 \frac{1 + \Gamma_g}{1 - \Gamma_g},$$
$$E = \frac{1 - \Gamma_g}{2} V_g.$$

The above equations together give

$$b_2=\frac{V_g}{2}\cdot\frac{S_{21}(1-\Gamma_g)}{1-\Gamma_g S_{11}-\Gamma_L S_{22}+\Gamma_g\Gamma_L(S_{11}S_{22}-S_{12}S_{21})}.$$

From the flow graph we see that $a_2=b_2\Gamma_L$. The total wave amplitude across the load is therefore

$$V_0= a_2+b_2=\frac{V_g}{2}\cdot\frac{S_{21}(1-\Gamma_g)(1+\Gamma_L)}{1-\Gamma_g S_{11}-\Gamma_L S_{22}+\Gamma_g\Gamma_L(S_{11}S_{22}-S_{12}S_{21})}.$$

Now the measured insertion ratio R_m is obtained by dividing the load voltage with network removed by the load voltage with network inserted. To remove the network, we set S_{11}, S_{22} equal to zero and S_{12}, S_{21} to unity. The result is

$$R_m=\frac{1 - S_{22}\Gamma_L-S_{11}\Gamma_g+\Gamma_g\Gamma_L(S_{11}S_{22}-S_{12}S_{21})}{(1-\Gamma_g\Gamma_L)S_{21}}.$$

If the source and load were reflectionless ($\Gamma_g=\Gamma_L=0$), the corresponding insertion ratio R_0 would be just

$$R_0 = \frac{1}{S_{21}}.$$

Hence, the quotient

$$Q=\frac{R_m}{R_0}=\frac{1-S_{22}\Gamma_L-S_{11}\Gamma_g+\Gamma_g\Gamma_L(S_{11}S_{22}-S_{12}S_{21})}{1-\Gamma_g\Gamma_L}$$

provides the measurement error due to network mistermination. In the common case where Γ_g and Γ_L are $\ll 1$, Q simplifies to

$$Q \sim 1 + \Delta$$

where Δ, the fractional error in nepers and radians, is given by

$$\Delta = -S_{11}\Gamma_g - S_{22}\Gamma_L + \Gamma_g\Gamma_L(1 + S_{11}S_{22} - S_{12}S_{21}).$$

For reciprocal structures, S_{12} is equal to S_{21}; these in turn are equal to the reciprocal of the design insertion ratio R_0.

As an example of the application of the expression for Δ, consider the measurement of a network having $|S_{11}| = |S_{22}| = 0.3$ (corresponding to a VSWR of 1.85) and $|S_{12}| = |S_{21}| = 1$. Then, if source and load were such that $|\Gamma_g| = |\Gamma_L| = 0.02$ (VSWR of 1.04), we could expect maximum errors of 0.11 db or 0.73 degrees, depending on the phases of the S's and Γ's.

DANIEL LEED
Bell Telephone Labs.
Murray Hill, N. J.

* Received by the PGMTT, June 27, 1961

[1] J. K. Hunton, "Analysis of microwave measurement techniques by means of signal flow graphs," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES vol. MTT-8, pp. 206–212; March, 1960.

Contributors

Edward E. Altshuler (S'54–M'55) was born in Boston, Mass., on January 10, 1931. He received the B.S. degree in physics from Northeastern University, Boston, in 1953, the M.S. degree in physics from Tufts University, Medford, Mass., in 1954, and the Ph.D. degree from Harvard University, Cambridge, Mass., in 1960.

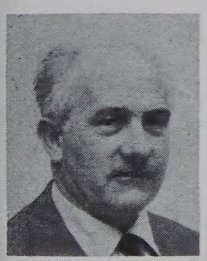
E. E. ALTSHULER

From 1954 to 1957, he was employed by Sylvania Electric Products, Inc., Waltham, Mass., as an antenna engineer. He is presently employed by the Electromagnetic Radiation Laboratory of Air Force Cambridge Research Laboratories, Bedford, Mass.

Dr. Altshuler is a member of Sigma Xi and Sigma Pi Sigma.

❖

Giorgio Barzilai (SM'50) was born in Rome, Italy, on June 23, 1911. He received the Ph.D. degree in industrial engineering from the University of Rome in 1935.



G. BARZILAI

From 1935 to 1948, he was at the Istituto Superiore P. T. Rome, Italy, and from 1938 to 1941 at the Italian National Council of Research, Rome, Italy. During this period, he was interested in the development of short wave antennas and transmitters. From 1941 to 1947, he was engaged in radar development in the Italian Air Force, followed by one year of antenna research at the University of Birmingham, England. From 1949 to 1954, he was at the Polytechnic Institute of Brooklyn, N. Y., engaged in teaching and microwave research activities. Since 1955, he has been with the University of Rome, where he is at present professor of applied electronics.

❖

Morris F. Bolster was born in Voorheesville, N. Y., on April 27, 1914. He worked for the General Electric Company in Schenectady, N. Y., as a test man and engineering assistant from 1941 to 1948. He received the B.S.E.E. and the M.S. degrees in mathematics from Union College, Schenectady, in 1951 and 1955, respectively.

M. F. BOLSTER

He has been employed by the General Engineering Laboratory of the General Electric Company in Schenectady since July, 1951. From then until 1953 he worked on mathematical analysis of engineering problems. Since 1953, he has been engaged in the development of microwave components.

❖

Seymour B. Cohn (S'41–A'44–M'46–SM'51–F'59), for a photograph and biography, please see p. 674 of the November, 1960, issue of these TRANSACTIONS.

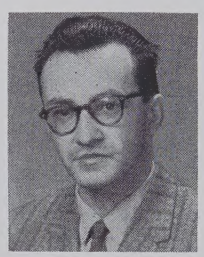
❖

Robert E. Collin (M'54–SM'60), for a photograph and biography, please see p. 208 of the March, 1961, issue of these TRANSACTIONS.

❖

Sverre T. Eng (M'58) for a photograph and biography, please see p. 105 of the January, 1961, issue of these TRANSACTIONS.

Wilhelm H. Eggimann was born in Zürich, Switzerland, in April, 1929. He received the diploma in electrical engineering in 1954 from the Swiss Federal Institute of Technology, Zürich.



W. E. EGGIMANN

From 1954 to 1956, he worked as an Instructor and Research Assistant at the same institute. He entered Case Institute of Technology, Cleveland, Ohio, in 1956, where he received the M.S. and Ph.D. degrees in electrical engineering in 1959 and 1961, respectively. In 1961, he joined the Professorial Staff of the Electrical Engineering Department at Case Institute. He is presently engaged in research work on ferrites in microwave applications and diffraction theory.

Dr. Eggimann is a member of Sigma Xi.

❖

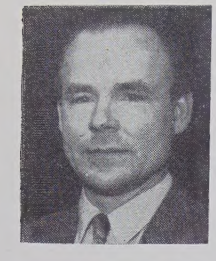
Giorgio Gerosa (A'58) was born in Rome, Italy, on August 21, 1931. He received the Ph.D. degree in electrical engineering from the University of Rome in 1956.

G. GEROSA

Since 1956, he has been at the University of Rome, engaged in microwave research and teaching activities as "professore incaricato" of microwaves. His primary interests are in propagation of electromagnetic waves in anisotropic media.

Dr. Gerosa is a member of the Associazione Elettrotecnica Italiana and the Société Française des Electroniciens et des Radioélectriciens.

A. E. Karbowiak was born in Warsaw, Poland, on March 1, 1923. He received the B.S. degree in 1950 and the Ph.D. degree in electrical engineering in 1954 from the University of London, England.



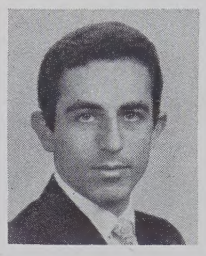
A. E. KARBOWIAK

From 1951 to 1954 he was at University College, London, carrying out research on surface waves. In 1954 he joined Standard Telecommunication Labs., where he was responsible for research on various aspects of long-distance communication by waveguide. Since 1958 he has been head of the Microwave Department.

Dr. Karbowiak is an associate member of the IEE.

❖

Nicholas J. Kolettis was born in Calimnos, Greece, on February 15, 1933. He received the B.S.E.E. and M.S.E.E. degrees in 1958 and 1961, respectively, from the Case Institute of Technology, Cleveland, Ohio.



N. J. KOLETTIS

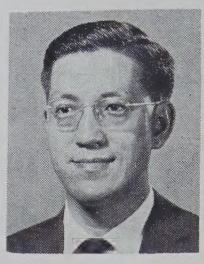
While attending graduate school, he was a Research Assistant with the Microwave Group of the Electrical Engineering Department at that Institute. At present, he is a Member of the Technical Staff of the RCA Laboratories, Princeton, N. J., where he is engaged in research work on microwave applications of solid state devices.

Mr. Kolettis is a member of Eta Kappa Nu and Tau Beta Pi, and an associate member of Sigma Xi.

❖

George L. Matthaei (S'49–A'52–M'57), for a photograph and biography, please see page 675 of November, 1960, issue of these TRANSACTIONS.

Charles C. H. Tang was born in Shanghai, China, on September 27, 1924. He received the B.S. degree in physics in 1946, from the University of Shanghai. He received the M.S. degree in electrical engineering in 1952, from Oklahoma State University, Stillwater, Okla. He entered Harvard University, Cambridge, Mass., on a Gordon McKay Scholarship, and received the Ph.D. in applied physics in 1956.



C. C. H. TANG

From 1946 to 1949, he taught in the physics department at the University of Shanghai. From 1950 to 1951, he was Chief Accountant at the Universal Textile Company, Ltd., Hongkong, China. From 1956 to 1957, he was at Harvard University as a Post-Doctoral Fellow. From 1957 to 1958, he was an Assistant Professor at the University of California, Berkeley, Calif. From 1958 to 1959, he was an Associate Professor in Physics at Tunghai University, Taiwan. Since 1959, he has been with the Bell Telephone Labs., Murray Hill, N. J.

Dr. Tang is a member of Sigma Xi and Phi Kappa Phi.

INSTITUTIONAL LISTINGS

The IRE Professional Group on Microwave Theory and Techniques is grateful for the assistance given by the firms listed below, and invites application for Institutional Listing from other firms interested in the Microwave field.

AIRBORNE INSTRUMENTS LABORATORY
A Division of Cutler-Hammer, Inc.
Deer Park, Long Island, New York

Research and Development

AIRTRON, INC.
A Division of Litton Industries
200 East Hanover Ave., Morris Plains, N.J.

Designers and Producers of Complete Line of Microwave Electronic and Aircraft Components

ALFORD MANUFACTURING COMPANY
299 Atlantic Ave., Boston 10, Mass.

RF Instruments, Coaxial Components, Antennas and Air Navigation Aids

CASCADE RESEARCH
Div. of Lewis & Kaufman Electronics Corp.
5245 San Fernando Rd., Los Angeles, Calif.

Research, Development, Production: Microwave Ferrite Devices, Microwave Components & Subsystems

EIMAC TUBES, EITEL-McCULLOUGH, INC.
301 Industrial Way, San Carlos, Calif.

Microwave Tubes, TWT-VTM-Reflex Klystrons, Power Grid Tubes, Amplifier Klystrons

ELECTRONIC SPECIALTY CO.
5121 San Fernando Rd., Los Angeles 39, Calif.

Airborne and Ground Antennas, Towers, Waveguides, Microwave Components, Complete Radiating Systems

FXR, INC.
25-26 50th Street, Woodside 77, N.Y.

Precision Microwave Test Equipment, High Power Microwave Electronics, Microwave Components & Instrumentation

ITT FEDERAL LABORATORIES
500 Washington Ave., Nutley 10, N.J.

Line-of-Sight and Over-the-Horizon Microwave Systems; Test Equipment and Components

LITTON INDUSTRIES
Electron Tube Division
960 Industrial Rd., San Carlos, Calif.

Magnetrons, Klystron, TWT's, Noise Sources, BWO's, Display Devices, CFA's, Switch Tubes, MM Wave Tubes

MELABS
3300 Hillview Ave., Palo Alto, Calif.

Microwave Components and Systems for Communications, Countermeasures, Reconnaissance, Radar, Etc.

MICROLAB
570 West Mount Pleasant Ave., Livingston, N.J.

Designers and Manufacturers of a Complete Line of Coaxial Microwave Components and Cavity Filters

MICROWAVE ASSOCIATES, INC.
South Avenue, Burlington, Mass.

Waveguide Devices, Microwave Diodes, Magnetrons, Duplexers, TWT's, Ferrite & Solid-State Devices

(See outside back cover for additional listings.)

The charge for an Institutional Listing is $50.00 per issue or $210.00 for six consecutive issues. Applications for Institutional Listings and checks (made out to The Institute of Radio Engineers, Inc.) should be sent to Robert A. Rivers, PGMTT Advertising Editor, Aircom Inc., 354 Main St., Winthrop 52, Mass.

INSTITUTIONAL LISTINGS

THE MICROWAVE JOURNAL
1330 Beacon St., Brookline 46, Mass.

A Magazine Devoted to the Interests of Engineers Working at Microwave Frequencies

STEWART ENGINEERING COMPANY
Santa Cruz, Calif.

Backward Wave Oscillators & Other TW Type Tubes; Controlled Atmosphere Furnaces & Precision Spot Welder

NARDA MICROWAVE CORP.
Plainview, L.I., New York

Microwave Test Equipment & Components, Modulators, Bolometer & Thermistors, Ferrite Devices

SYLVANIA MICROWAVE DEVICE OPERATIONS
Sylvania Electric Products Inc.
500 Evelyn Ave., Mountain View, Calif.

Magnetrons, Klystron, TWT's, BWO's, Ferrite Devices, Waveguide Windows, Microwave Diodes

PHELPS DODGE COPPER PRODUCTS CORP.
300 Park Ave., New York 22, N.Y.

Styroflex, Foamflex Air Dielectric Aluminum Sheathed Coaxial Cables, Connectors and Waveguides

TECHNICRAFT DIVISION, Electronic Specialty Co.
Thomaston, Conn.

Designers, Developers and Producers of Microwave Components, Transmission Lines, and Assemblies

PRD ELECTRONICS, INC.
202 Tillary St., Brooklyn 1, N.Y.

Complete Line of Microwave and Electronic Test Equipment Waveguide and Coaxial Components

WATKINS-JOHNSON COMPANY
3333 Hillview Ave., Palo Alto, Calif.

Res., Dev., Mfg. Microwave Electron Devices, TWT's, BWO's, Parametric Amplifiers, Microwave Systems

SAGE LABORATORIES, INC.
3 Huron Dr., Natick, Mass.

Microwave Attenuators, Couplers, Crystal Holders, Filters, Hybrids, Mixers, Rotary Joints

WEINSCHEL ENGINEERING COMPANY, INC.
10503 Metropolitan Ave., Kensington, Md.

Attenuation Standards, Coax Attenuators, Insertion Loss Test Sets, Voltage & Power Calibrators

SCIENTIFIC-ATLANTA, INC.
2162 Piedmont Rd., N.E., Atlanta 19, Ga.

Complete Antenna Pattern Range Instrumentation, Special RF and Antenna Systems Development

WHEELER LABORATORIES, INC.
Great Neck, N.Y.
Antenna Lab., Smithtown, N.Y.

Consulting Services, Research & Development, Microwave Antennas & Waveguide Components

(See inside back cover for additional listings.)